(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number
WO 02/34939 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/CA01/01487

(22) International Filing Date: 22 October 2001 (22.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/695,545 23 October 2000 (23.10.2000) US

(71) Applicant: MBI FERMENTAS INC. [CA/CA]; 830 Harrington Court, Suite 3, Burlington, Ontario L7N 3N4 (CA).

(72) Inventor: YU, Alex, H., C.; Block 1E, #04-42, Gillman Heights, 104001 Singapore (SG).

(74) Agent: SMART & BIGGAR; Attn: KANG, Yoon (Ms.), 438 University Avenue, Box 111, Suite 1500, Toronto, Ontario M5G 2K8 (CA).

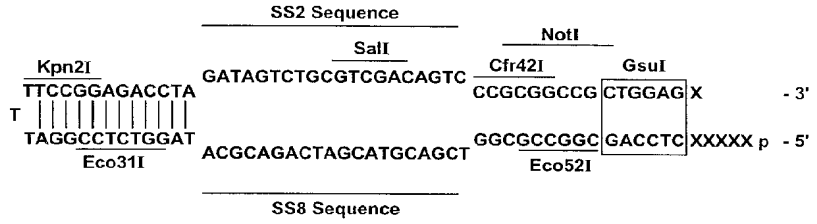
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

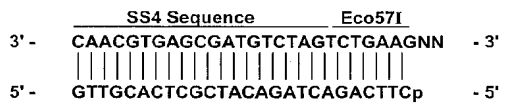
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL METHOD OF GENOMIC ANALYSIS



Where Xs represent the respective overhangs of restriction endonucleases

1a



1b

(57) Abstract: The present invention provides a novel method of genomic analysis. In the method, an extension primer generated by capturing an unknown DNA sequence with an adaptor containing a class IIS RE recognition site is used in an extension/amplification reaction with a DNA sample to amplify a unique product containing the captured sequence. The method of the invention therefore allows identification and analysis of unknown sequences upstream to a known region or a sequence. The invention has a particular application to genotyping, including bacterial strain typing. Specifically, the invention provides a method of preparing extension primers to capture sequences from different genomic locations and their use as genetic markers. The invention also provides adaptors to prepare extension primers.

WO 02/34939 A2

NOVEL METHOD OF GENOMIC ANALYSIS

Field of The Invention

The present invention relates to the field of genomic analysis. More specifically, it relates to a process for generating extension primers by capturing sequences from DNA fragments and their use in genomic analysis, including in bacterial genotyping.

Background of the Invention

During outbreaks of bacterial infections in hospitals, strain typing is often required as an integral part of epidemiological investigations. The data generated from strain typing is used to identify outbreak-related strains and to distinguish epidemic from endemic or sporadic isolates. With the introduction of molecular strain typing methods in the past 5 years, traditional methods such as bacteriophage typing and serotyping, have been supplemented or replaced in many laboratories. This is driven by the need for a strain typing method that can be used to type a broader array of bacterial species. Although bacteriophage typing and serotyping work very well in epidemiological studies of *Staphylococcus aureus* and *Salmonella* species, they are limited in their applications to different bacterial species. In addition to a broader application to a variety of bacterial species, the newer molecular methods can better distinguish the strains and, in some cases, are less time consuming. The most commonly used molecular strain typing methods in clinical and reference laboratories are pulsed-field gel electrophoresis (PFGE), and random amplified polymorphic DNA (RAPD). These methods are based on DNA polymorphism either in the restriction patterns based on restriction digestion, or in the banding patterns of PCR amplification products and do not require the prior knowledge of the type of polymorphism, or the sequence information of the DNA fragments in the banding patterns.

PFGE analysis is essentially an improved method over the previously used method based on restriction fragment length polymorphism (RFLP). In the prior method, genomic DNA is digested by a frequently cutting restriction endonuclease (RE) into hundreds of fragments. These fragments are separated by agarose gel electrophoresis and then

transferred to a membrane. The DNA on the membrane is then hybridized with a specific DNA probe. The probe is so designed that only few fragments will hybridize to it. Variations in the number and size of the fragments detected by hybridization can be used to distinguish different bacterial strains. Ribotyping, the use of ribosomal RNA probe in RFLP, is one common typing method based on RFLP. But the enthusiasm for this system has diminished because it has been proven that this approach is only moderately discriminatory. The use DNA insertion element IS6110 is currently the method of choice for typing isolates of *Mycobacterium tuberculosis* based on RFLP. An insertion element is a DNA fragment with a defined structure that is capable of moving independently and inserting in multiple locations in plasmids or chromosome. Since the insertion element is mobile, the number and the locations of the insertions vary greatly from strain to strain. This approach has been proven to be reliable and discriminatory for isolates with more than 5 copies of IS6110. However, the approach is still limited to this particular species.

Because RFLP analysis is laborious to carry out, PFGE became the method of choice in typing strains based on fragment patterns produced by cleaving chromosomal DNA with restriction endonucleases. With the introduction of PFGE in 1984, it is now possible to separate large DNA fragments ranging from 10 to 1000 Kbp. In PFGE analysis, the chromosomal DNA is cut with a restriction endonuclease that generates approximately 10 to 30 fragments. PFGE is used to resolve all of these bands and the gel pattern generated can be used to type different bacterial strains. It is currently the method of choice for typing most species. Almost all bacterial species are typeable by PFGE analysis except *C difficile,* the DNA of which spontaneously degrades during cell lysis. Two major difficulties associated with PFGE are the technical demands of the procedure, and the initial cost of the equipment. Preparation of suitable genomic DNA requires 1 to 3 days, and another 3-4 days are required to run PFGE. The equipment costs about $10,000-$20,000. Therefore, many laboratories are still not equipped to carry out PFGE analysis.

Since the development of PCR, PCR based methods have been developed and applied to identify bacterial strains and species. In some applications, specific PCR primers are used to amplify unique products from the bacterial DNA for diagnostic purposes. In contrast, multiple arbitrary amplicon profiling (MAAP) techniques use non-specific primers

to target multiple and arbitrary amplicons as described by Caetano-Anolles, PCR Methods Applic. 3:85-94 (1993). The variation in number and length of the amplified DNA fragments can be used to determine the relatedness of bacterial strains. A number of variations of this approach has been developed. Random amplified polymorphic DNA (RAPD) analysis, arbitrarily primed PCR (AP-PCR), and DNA amplification fingerprinting (DAF) are MAAP techniques that differ in primer length, amplification stringency, and procedure used to resolve DNA patterns. The DNA profiles obtained from these methods are distinct due to the difference in the sizes of the primers and the amplification conditions employed. For example, RAPD analysis resolves fewer amplification products than DAF.

Since amplification is conducted at low annealing temperature during RAPD analysis, the amplification products obtained are greatly affected by the parameters of PCR. See Tyler, et al., J. Clin. Microbiol. 35:339-346 (1997). The reproducibility and discriminatory power of MAAP techniques have been under intensive investigation. Studies have shown that these methods have some value in rapidly discriminating between individual isolates but they are very susceptible to technical variations. It has been found that there can be substantial variations in the efficiency with which the primers initiate DNA synthesis at a particular site, depending on even the slight variations in the pH or ionic strength of the buffer used. The number of copies generated from a particular locus can be appreciably different on independent amplification of the same strain. Hence, it is difficult to obtain reproducible patterns from a single isolate tested on different days. In a multicenter study, it has been shown that inter-laboratory comparison of DNA patterns generated from RAPD is problematic.

RAPD is a random indeterminate amplification process that uses stochastic annealing events to randomly amplify DNA fragments. As an alternative to the arbitrary approach, known repetitive DNA motifs and highly conserved intergenic repetitive sequences have been exploited as primer binding sites for the amplification of bacterial DNA. Polymorphic DNA fingerprints can be obtained with such an approach, such as REP-PCR described by Versalovic, et al. Methods Mol. Cell Biol. 5:25-40 (1994). REP-PCR is a determinate process that targets multiple sites of known sequence but unknown locations, and amplifies DNA fragments between repetitive elements only. These repetitive sequences include the 33-

to 40-bp repetitive extragenic palindromic (REP) elements discovered in the genomes of *E. coli* and *S. typhimurium*, the 124- to 127 bp enterobacterial repetitive intergenic consensus (ERIC) elements discovered in gram-negative bacteria species, and the 154-bp BOX elements discovered in the genome of *S. pneumoniae*. All these motifs are genetically stable and differ only in their copy numbers and chromosomal locations between species. Their copy numbers vary from approximately 25 copies for BOX, approximately 30 to 150 copies for ERIC, to approximately 500 to 1000 copies for REP. Since only primers that bind to the repetitive motif are used in REP-PCR to obtain amplified products, adjacent copies of the repetitive motif have to be located close together, within the amplification range of PCR. This limits the use of REP-PCR to bacteria species with a high copy number of the repetitive motifs. Although REP-PCR gives more reproducible results, the discriminatory power is compromised due to the high stringency conditions used in PCR. This is because the sequences of the repetitive motifs are not 100% conserved. While consensus sequences can be derived and used as the primer sequence, they may not be applicable to all bacteria species.

Therefore, there is still a need for a rapid typing method that is reproducible and easy to perform.

To make a PCR-based typing method more reproducible, specific primers should be used with stringent PCR conditions. To make the typing method more discriminatory, primers that are known to amplify polymorphic region of the bacterial genome among strains should be used. In order to generate primers that are sequence-specific and that can amplify polymorphic DNA regions, polymorphic regions of different strains have to be identified and characterized. Therefore, there is the need to be able to easily and quickly identify multiple polymorphic sites in bacteria genomes. In the past, researchers have identified unique regions in a particular bacterial species that they can use to make probes or primers to identify specific strains. But the methodologies employed are laborious and time-consuming, and usually require massive cloning and screening efforts. It is impractical to use the same methodologies to discover multiple polymorphic loci for a large number of strains of the same bacteria species.

Summary of The Invention

In one embodiment, the invention relates to a method of analyzing DNA which includes the step of ligating a first oligonucleotide and a DNA fragment to form a ligated product. The oligonucleotide comprises a known sequence and a recognition site of a class IIS RE. The ligated product is then digested with the class IIS RE to release a second oligonucleotide comprising the first oligonucleotide and a sequence from the DNA fragment. A reaction mixture is formed comprising the DNA sample, the second oligonucleotide, a first amplification primer and a second amplification primer under conditions suitable to drive extension and amplification. The first amplification primer specifically hybridizes to the known sequence and the second amplification primer specifically hybridizes to a sequence in the DNA fragment. The presence of an extension product in the reaction mixture is then determined.

The invention also relates to a method of preparing an extension primer with captured genomic DNA sequences. The method includes the steps of digesting genomic DNA of a species to generate DNA fragments and ligating the resulting DNA fragments with a first oligonucleotide to form ligated products. The oligonucleotide comprises a known sequence and a recognition site for a class IIS RE. One or more ligated products are then amplified with a first and second primer. The first primer specifically hybridizes to the known sequence and the second primer specifically hybridizes to a sequence in the genomic DNA. The one or more amplified product is then digested with the class IIS RE to generate an extension primer comprising the first oligonucleotide and a sequence from the genomic DNA.

The invention further relates to a method of typing genomic DNA which includes the step of forming a reaction mixture comprising the genomic DNA, an extension primer, a first amplification primer and a second amplification primer under conditions suitable to drive extension and amplification. The extension primer comprises a known sequence and a first sequence of genomic DNA from another member of the species, the first amplification primer specifically hybridizes to the known sequence and the second amplification primer specifically hybridizes to a second sequence downstream from the first sequence of the genomic DNA of the other member of the species. The presence of an extension product in the reaction mixture is then determined.

The invention also relates to a DNA adaptor which may be used in the preparation of extension primers according to the invention. The adaptor according to the invention is a single stranded oligonucleotide capable of forming a partially double stranded structure having a closed end and an open end, the structure comprising at the closed end a first double stranded region, at the open end, a second double stranded region, between the first and second double stranded regions, a region of non-complementary sequences. The non-complementary sequences comprise a known sequence to which a primer can specifically hybridize, and the second double stranded region comprises a class IIS RE recognition site.

The adaptor of the invention may be included in a kit which is also contemplated by the invention.

The invention also relates to primers generated according to the invention, including primers of SEQ ID N0. 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO. 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 and SEQ ID NO: 27

Brief Description of the Drawings

Figure 1(a) and 1(b) illustrate the structures of a Sequence-Capture (SC) Adaptor and an amplification adaptor, respectively. The oligonucleotides that have their 5' ends phosphorylated were shown with a **p** at their ends.

Figure 2 is a schematic representation of amplification of unique DNA fragments by (1) generation of A-T-P, (2) A-T-P amplification, and (3) extension/PCR, in accordance with an exemplary embodiment of the present invention.

Figure 3 is a schematic representation of amplification and reversal of the functional directionality of A-T-P, in accordance with an exemplary embodiment of the present invention.

Figure 4 is a schematic representation of amplification of DNA fragments from an *in vitro* bacterial genomic library using an anchor primer, in accordance with an exemplary embodiment of the invention.

Figure 5 is a schematic representation of preparation of A-T-P from an amplified product of Figure 4, in accordance with an exemplary embodiment of the invention.

Figure 6 illustrates a portion of known map of E. coli dnak gene, illustrating the positions of the Pst I restriction sites and the PCR primer binding sites. Figure 6 also illustrates the amplified products obtained from extension/PCR reactions with A-T-P_{DNAK}, PSS2 and 3'DNAK, and with "reversed" A-T-P_{DNAK}, PSS4 and 5'DNAK2.

Figure 7(a) is a summary of the PCR Products obtained from *in vitro* libraries (PagI and EcoRI) of *S. aureus* strains with respect to known map of mec DNA regions. Figure 7(b) is a summary of the PCR Products obtained from *in vitro* libraries (HindIII and PstI) of *S. aureus* strains with respect to known map of mec DNA regions.

Figure 8 is a summary of the PCR Products obtained from *in vitro* library (EcoRI) of *S. aureus* strains with respect to known map of pSK1 DNA.

Figure 9 is a summary of the PCR Products obtained from *in vitro* libraries (PagI, HindIII and EcoRI) of *S. aureus* strains with respect to known map of pSK41 DNA.

Detailed Description of the Invention

The following definitions are provided to facilitate an understanding of the present invention.

The term "class IIS restriction endonuclease (RE)" is used to describe a subgroup of class II RE that cleaves DNA at precise distances outside their recognition sites. Unlike class II RE that cut within the recognition site, class IIS RE cuts DNA but leaves the recognition site intact. Depending on the individual class IIS RE, the precise distance between the cleavage site and the recognition site, and the type of ends generated vary.

The term "primer" is used to describe single-stranded (ss) oligonucleotide that can be used to initiate DNA synthesis after hybridizing to its complementary DNA sequence in the presence of DNA polymerase and dNTP. It can also be used together with another primer in PCR reactions to amplify DNA fragments.

The term "extension" is used to describe the process of DNA synthesis initiated by an oligonucleotide primer. Primer with "sufficient" complementary sequence to a DNA strand at its 3' end can hybridize to the DNA strand and has its length extended along the complementary DNA strand in the presence of DNA polymerase and dNTP. "Sufficient" is used to mean the primer is thermodynamically stable to remain bound to the DNA template strand, such that the DNA polymerase is able to extend the primer by copying the template strand.

The term "sequence-capturing (SC) adaptor" is used to describe an oligonucleotide that has complementary sequences to allow the formation of a partially double stranded structure which folds itself into a stem-loop structure as exemplified in Figure 1(a). The open ends of these adaptors can be blunt, with a 5' overhang, or with a 3' overhang. In the non-complementary region of these adaptors, a unique sequence is present for PCR amplification. The unique sequence is of a length sufficient for specific hybridization by a primer. The uniqueness of this sequence is defined by the statistical probability of occurrence of a random sequence of that particular length. This unique sequence may or may not contain a specific sites recognized by a restriction endonuclease. One or more restriction sites recognized by a class IIS restriction endonuclease is present in the adaptor.

The term "captured sequence" is used to describe the DNA sequence that is joined to the SC adaptor following ligation of the SC adaptor to a DNA fragment and digestion with a class IIS restriction endonuclease. The length of the captured sequence is dependent on the class IIS restriction endonuclease used to cut the adaptor from the DNA fragment.

The term "adaptor-turned-primer" or "A-T-P" is used to describe a SC adaptor with captured sequence. The end of the A-T-P that is separated from the DNA fragment can be blunt, sticky end with 5' overhang, or sticky end with 3' overhang. A sticky end will

facilitate ligation to an "amplification adaptor" described below. A-T-P can be used as primer to initiate DNA synthesis.

The term "functional directionality" of A-T-P is used to describe the extension direction of the A-T-P. In DNA synthesis, the direction of the synthesis is from the 5' end to the 3' end. When a primer hybridizes to the DNA template, only its 3' end can be extended. When using A-T-P for extension, only the 3' end of the upper strand of the captured sequence can be extended. The opposite strand of the captured sequence in the A-T-P is incapable of initiating primer extension. Therefore, only one strand of the A-T-P entity is functional as an extension primer and it determines the direction of the extension along the DNA template.

The term "amplification adaptor" is used to describe an entity formed from 2 oligonucleotides that have significant complementary sequences to allow the formation of a double-stranded (ds) region as exemplified in Figure 1(b). One end of the adaptor is blunt. At this end, a unique sequence is present for PCR amplification. This sequence has a length sufficient for specific hybridization by a primer. This unique sequence may or may not contain specific sites recognized by a restriction endonuclease. At the other end, a restriction site recognized by a class IIS restriction endonuclease is present. The class IIS restriction endonuclease preferably has the same cutting pattern (*i.e.* cuts at the same distance away from the recognition site) as the class IIS restriction endonuclease used to cut the SC adaptor. This end of the amplification adaptor can be blunt, with a 5' overhang or with a 3' overhand. In one illustrative embodiment, a sticky end of a random sequence is a 2-bases 3' overhang and represents 16 (4^2) possible combinations of ends. The amplification adaptor is used to reproduce A-T-P or to reverse the functional directionality of A-T-P.

The term "*in vitro* library" is used to describe genomic DNA that has been digested with an RE, and then ligated to SC adaptors. Genomic DNA is digested with different REs and then ligated to SC adaptors to generate different *in vitro* libraries.

The present invention provides a method, referred to generally as SCOT (sequence capturing oligonucleotide technique) for preparing extension primers from unknown DNA sequences. Following ligation of a defined oligonucleotide to unknown DNA, the ligated

product is digested with class IIS RE that recognizes the enzyme site on the oligonucleotide. Some of the DNA sequences are thereby captured and can be used directly as sequence specific extension primer. A specially designed SC adaptor is provided to facilitate the capture of sequence of complex DNA .

Provided a sequence 3' downstream to the unknown DNA sequences is known, the extension primer can be used in an extension/amplification reaction to specifically extend the captured sequence and amplify the upstream sequences to the known 3' sequence or region. The upstream sequence so amplified may then be sequenced or subject to other genomic analysis. The extension/amplification reaction does not require prior sequence information of the captured sequence. This is in contrast to U.S patent no. 5,695,937 which discloses the use of a linker containing a class IIS RE site to capture expressed sequence and formation of concatermers of the captured sequence. The captured sequence then must be cloned and sequenced to provide quantitative and qualitative data about gene expression.

The extension primer of the invention may be amplified, and its functional directionality may be reversed.

The extension primer generated by the present invention has a particular application in identification of polymorphic regions and therefore, to genotyping. In particuar, by preparing extension primers which have captured sequences from different genomic locations, these extension primers can identity regions of polymorphism and thereby serve as genetic markers and/or provide the necessary sequence information to design such markers. While this aspect of the invention is exemplified below with respect to bacterial DNA strain typing, it will be readily appreciated that the invention extends equally to less complex genome and to other genome whose complexity is of the order of bacterial genome, including certain human cDNA or library clones of human DNA.

The present invention may be carried in accordance with the teachings that follow using standard techniques of molecular biology well known to persons skilled in the art.

(1) Sequence Capturing

Sequence capturing is accomplished by ligating a SC adaptor to a DNA fragment with complementary sticky ends or with blunt ends. It is based on the cleavage principle of class IIS RE that cuts away from its DNA recognition site. Since the SC adaptor includes a class IIS restriction site, the class IIS RE is used to cut DNA ligated to the SC adaptor. The released SC adaptor now "captures" some DNA sequence from the DNA fragment that it ligated to. The number of bp captured is dependent on the class IIS RE used; it generally ranges from 1 to 16 bp with variable sizes of overhangs. The class IIS RE recognition site can be positioned at the open end of the adaptor so as to maximize the length of the captured sequence.

The captured sequence should be of a length that enables sequence specific primer extension. That length will depend on the complexity of the DNA that is subject to analysis. The more complex the DNA, the more captured sequence is required. Preferably the length of the captured sequence is at least 12 bp long. More preferably, the captured sequence is 16 bp long. This length of captured sequence is sufficient to serve as a sequence-specific extension primer for human genomic DNA.

While the inventor has observed that the stem-loop structure of the SC adaptor is necessary when ligating to highly complex DNA populations like RE digested bacterial genomic DNA to avoid non-specific background amplification, it will be appreciated that for an isolated DNA fragment and other simple DNA populations, any oligonucleotide having a unique sequence for PCR amplification and a class IIS RE recognition site may be used in place of the SC adaptor.

In an exemplary embodiment, SC adaptor with a 4-bases 3' overhang is ligated to a DNA fragment with the complementary end. After ligation, the ligation reaction may be purified for example with a DNA extraction kit (Fermentas, Lithuania). The DNA extraction kit allows for the recovery of long DNA fragments by reversible binding of DNA to glass particles. The unligated adaptors are removed in this process. The purified fragment/adaptor is then digested with a class IIS RE which recognizes the class IIS RE site present in the SC adaptor. The released SC adaptor contains some DNA sequence from the DNA fragment at its 3' end with respect to the top strand, and at its 5'end with respect to

the bottom strand. As discussed below, the A-T-P that results may be amplified, or it may be reversed with respect to its functional directionality, before its use in an extension/PCR reaction.

The DNA fragment used in the ligation could be a gel-purified fragment from a RE digestion, a DNA product of a PCR reaction, RE-digested plasmid, or RE-digested genomic DNA fragments. The DNA may or may not be dephosphorylated depending on the SC adaptor to be used for ligation. If multiple DNA fragments are used in the ligation, the A-T-P generated will contain different captured sequences. This A-T-P mixture can be used in an extension/PCR reaction but the number of different A-T-P's (*i.e.* different captured sequences) should be limited so as to generate unique amplification products following extension/PCR reaction. Preferably the number of different A-T-P's is less than 20.

When DNA fragments of high complexity (bacterial genomic DNA) is used in the ligation, a specific DNA fragment from the ligation mixture is amplified by PCR as discussed below. The PCR product may then be purified using agarose gel electrophoresis before subjecting to the class IIS RE digestion to generate A-T-P.

If necessary, A-T-P can be partially purified, for example, with molecular weight cut-off filters. By passing the class IIS RE digestion of fragment/adaptor through a 100,000 MW cut-off filter in a spin unit (Gilson), 80-90% of A-T-P can be separated from long DNA fragments and recovered.

(2) Amplification of A-T-P

ATP can be reproduced in large quantities by ligating to an amplification adaptor and amplifying the ligated product. For example, an A-T-P generated by cutting with a class IIS RE, having a 2-bases 3' overhang can be ligated to an amplification adaptor with 2-bases 3' overhangs of random sequence ($4^2 = 16$ combinations) and the ligated product PCR amplified.

After ligation to the amplification adaptor, A-T-P is amplified in 10 PCR cycles to prepare a stock solution using primers that are complemenary to the known unique sequences on the SC adaptor and on the amplification adaptor. Subsequent working solution

of A-T-P is prepared from a 20-cycles PCR amplification of a small aliquot of the stock A-T-P solution using a biotinylated amplification adaptor primer and the SC adaptor primer. These PCR reactions can be carried out under the standard temperature cycling profiles used by those skilled in the art. The duration of the temperature cycles in the PCR reactions should be modified according to the size of the amplified product. As the rates of polymerization by DNA polymerases are known, the appropriate time required to obtain a full length amplified product can be determined by those skilled in the art.

To release the A-T-P, the PCR product is digested with the class IIS RE that recognizes the restriction site on the SC adaptor. The A-T-P is separated from the amplification adpator using streptavidin-immobilized beads. The amplified A-T-P is functional in sequence-specific primer extension.

Preferably, the amplified A-T-P/adaptor complex is first bound to the streptavidin beads. The A-T-P is then released from the beads by digesting with a class IIS RE that recognizes the restriction site on the SC adaptor. The supernatant containing the A-T-P is separated from the beads after a brief centrifugation step.

Affinity Tags other than biotin on the amplification adaptor primer can be used to separate ATP from the amplification adaptor after class IIS RE digestion.

(3) Extension/PCR amplification

A-T-P can be used to amplify a unique DNA fragment in an extension/PCR reaction. For efficiency, the reaction is preferably carried out as a 2 step extension/PCR reaction. In the extension step, A-T-P is added to a DNA mixture (e.g. genomic DNA) and the reaction is heat denatured. After the temperature of the reaction is brought down to 37°C, the A-T-P hybridizes to its complementary genomic DNA sequence and its sequence is extended with the addition of a DNA polymerase in the presence of dNTP. The newly synthesized DNA strand now has the unique sequence of the SC adaptor at its 5' end. PCR reaction is then carried out with a primer complementary to the unique sequence of the SC adaptor, and another primer that is specific to a region 3' downstream from the captured sequence.

In one illustrative embodiment, the extension reaction is carried out with Klenow fragment in a PCR buffer. The extension reaction is carried out at 37°C for 30 min. At the end of the extension reaction, PCR primers and Taq DNA polymerase are added and the reaction is subjected to temperature cycling as in a standard PCR reaction.

The extension reaction may also be carried out with Taq DNA polymerase. The extension reaction is carried out at 37°C for 30 seconds and then at 72°C for 2 minutes. The extension reaction may be repeated, preferably 5 times. At the end of the extension, PCR primers are added and the reaction is subjected to the standard PCR reaction conditions.

After the extension/PCR reaction, a specific DNA fragment is obtained comprising the SC adaptor sequence at the 5' end of the fragment. The length of this fragment is determined by the distance between the captured sequence on the A-T-P and the 3' PCR primer.

A-T-P can be regenerated from this fragment by cutting with the class IIS RE. If the A-T-P is functionally reversed as described below, it can be used to primer extend in the opposite direction.

(4) Reversal of functional directionality

A-T-P generated is a double stranded (ds) DNA molecule with either a blunt end or a sticky end. The captured sequence is at the 3' end with respect to the top strand and 5' end with respect to the bottom strand of the A-T-P. Since primer extension is only carried out in the 5' to 3' direction, only the top strand can be used as sequence-specific primer with respect to the captured sequence.

The class IIS RE, which recognizes the restriction site on the amplification adaptor, upon cleavage should preserve the length of the captured sequence at least to the extent necessary to enable sequence specific extension. Preferably it has the same cutting pattern as the class IIS RE used to cleave SC adaptor. After ligating the amplification adaptor to A-T-P, the A-T-P/ amplification adaptor complex can be amplified in a 10-cycles PCR

reaction as described for the amplification of A-T-P. Subsequent working solution for preparing reversed A-T-P may be prepared from a 20-cycles PCR amplification of a small aliquot of the stock A-T-P solution using a biotinylated SC adaptor primer and the amplification adaptor primer.

The amplification reaction is digested with the class IIS RE that recognize the RE site on the amplification adaptor. The captured sequence is now transferred to the amplification adaptor from A-T-P. The bottom strand of the captured sequence is now at the 3' end of the amplification adaptor, and can be used in primer extension as sequence-specific primer. See Figure 3.

The amplified A-T-P/adaptor complex may be first bound to the streptavidin beads. The reversed A-T-P is then released from the beads by digesting with a class IIS RE that recognizes the restriction site on the amplification adaptor. The supernatant containing the reversed A-T-P is separated from the beads after a brief centrifugation step. The reversed A-T-P is functional in sequence-specific primer extension. Its functional (priming) direction is opposite to that of the A-T-P from which it was prepared.

Bacterial Strain Typing

(1) Creation of *In Vitro* Libraries

Traditional PCR amplification of a DNA fragment requires two primers that are complementary to the sequence of opposite strands of the DNA fragment and that are some distance apart in their binding sites. Therefore, the DNA sequence information must be available in order to synthesize the primers. In the present invention, only one PCR primer complementary to a known DNA region is required to initiate the amplification process. In conjunction with an "*in vitro* library", specific DNA fragments around a known DNA region can be amplified with 1 target-specific primer hereinafter referred to as the "anchor primer".

With reference to Figure 4, to create an *in vitro* library of a particular restriction endonuclease, purified bacterial genomic DNA is digested with a restriction endonuclease. Following digestion, the genomic DNA may be purified. The digested DNA is then ligated

to the SC adaptor. The ligation reaction may be subjected to PCR amplification without any further purification. See Figure 4.

The genomic DNA may be digested with any RE that generate cleaved fragments of a length that can be amplified. As such, the selection of the RE will depend on the DNA polymerase to be used to amplify the DNA fragment ligated to the SC adaptor.

In one illustrative embodiment, multiple *in vitro* bacterial genomic libraries are created using restriction endonucleases that cut at a frequency of once per 2000 - 8000 bp in the bacterial genome. These restriction endonucleases produce 5' or 3' overhangs.

(2) Amplification of Multiple Fragments from the *In Vitro* Genomic Libraries

An anchor primer is required together with a primer specific to the adaptor (adaptor primer) to amplify DNA fragments from the *in vitro* genomic libraries. In essence, a DNA fragment is amplified between a restriction site (the one used for creating the *in vitro* library) and the anchor sequence. To identify polymorphic regions, the anchor primer should specifically hybridize to variable positions within the bacterial genome to allow amplification of different DNA fragments from different regions of the bacterial genomes. Candidates for the anchor sequence include insertion sequences and repetitive elements. The insertion sequences and repetitive elements are semi-conserved sequences that are present in multiple copies and in various locations of bacterial genomes.

In one illustrative embodiment, an insertion element is chosen as the target for the anchor primer. By using a particular RE-digested *in vitro* library in PCR, DNA fragments between the insertion element and the RE site is amplified. Polymorphism (between the insertion elements and the restriction sites in front of them) in terms of different lengths and/or absence of the amplified products may be observed among different bacterial strains. If more than one copy of the insertion element is present in the bacterial genome, then more than one band should be obtained unless the RE site is too far away from the IS element. In the example described below, a PCR primer with sequence complementary to an internal region of the transposase gene in the IS431 insertion element was prepared as the anchor primer.

The ligation reactions are amplified with the adaptor primer and the anchor primer in PCR reactions. Any method of analysis to detect the amplification product and its size can be employed to analyze the amplified products. Agarose gel electrophoresis provides a convenient method for such analysis.

(3) Preparation of the SCOT Adaptor-turned-Primer (A-T-P) from Amplified Products

The preparation steps of A-T-P are shown in Figure 5. After obtaining the amplified products from the PCR reactions of *in vitro* libraries, one or more bands not common to, and therefore capable of differentiating different strains or species is purified from a gel slice cut from the agarose gel. The purified band is digested with the class IIS RE that recognizes the site on the adpator, to release the adaptor along with the captured sequence (A-T-P).

The A-T-P can then be ligated to an amplification adaptor. The ligated A-T-P/adaptor complex is amplified in 10 cycles of PCR. This PCR reaction can become the stock for that particular A-T-P and an aliquot of this stock PCR may be used to reproduce more A-T-P, with the adaptor primer and a biotinylated amplification adaptor primer, in 20 cycles of PCR. The biotinylated product is then run in a 2.5% agarose gel or other appropriate concentration to separate it from any large contaminating fragments. The short amplified product is cut from the gel and further purified with streptavidin beads. The bound amplified products is digested with the class IIS RE that recognizes the site on the adpator to release the A-T-P. The supernatant of the digestion is the working solution of A-T-P to be used in the extension/PCR reaction.

(4) Extension/PCR of the A-T-P on Different Bacterial Strains to Identify Polymorphism

A-T-Ps are prepared from amplified products obtained from different *in vitro* libraries of different bacterial strains or species. As discussed above, each of the A-T-P includes the adaptor primer sequence and the captured sequence from the original bacterial DNA that the adaptor ligated to. A-T-Ps are used in extension/PCR reactions to amplify unique fragments from bacterial genomic DNA.

In one illustrative embodiment, A-T-P is mixed with undigested bacterial genomic DNA of unidentified strain or species in the presence of Taq DNA polymerase. After denaturation, the upper strand of the A-T-P should be able to prime and extend on the genomic DNA if the complementary sequence of the captured sequence is present. After five repeat extension cycle, the reaction is mixed with the anchor primer and the adaptor primer. Exponential PCR is then allowed to proceed for 30 cycles. Amplified products would only be obtained if the captured sequence and the anchor sequence were located sufficiently close in the bacterial DNA.

The subject method therefore can be used to construct nucleic acid "fingerprints" of different bacteria strains. Such fingerprints are specific to bacteria strains, and can be used to identify the type of bacteria in infections outbreaks. For example, A-T-Ps can be prepared from *in vitro* libraries of multiple strains of a particular species. Using the A-T-Ps in extension/PCR reactions of genomic DNA of the strains, a set of A-T-Ps that could distinguish all of the strains by the amplification products could be identified. By using this set of A-T-Ps, a unique DNA "fingerprints" will be obtained for each strain.

The amplified products of the extension/PCR reactions may also be sequenced. Optimized PCR primers can then be designed around the captured sequence and are within the scope of the invention. They may be more effective than the original A-T-Ps in typing bacteria. The extension/PCR reaction with A-T-P and 2 PCR primers in most cases is expected to be less sensitive than a PCR reaction using a pair of PCR primers. It may be useful to develop optimized PCR primers from amplified products for highly sensitive strain typing methods. A database of PCR primers may be created to facilitate the development of a multiplex PCR approach to strain typing. Depending on the anchor primer used, the database could be applicable to a group of bacterial species.

The subject method therefore can also be used to identify and prepare DNA probes for a particular strain or closely related strains. If the amplified fragment of the extension/PCR reaction is present in the genome of only 1 strain, it can be used as strain specific probe. The amplified fragment can be made into a radioactive probe easily by including a radioactive nucleotide in the extension/PCR reaction. Such modified primers are also within the scope of the invention.

This invention further contemplates a kit comprising one or more reagents of the invention. The kit may include, in packaged form, a multicontainer unit having multiple SC adaptors of various sticky ends; an adaptor primer; an anchor primer; an amplification adaptor; an amplification adaptor primer; each of four different nucleoside triphosphates; restriction endonucleases that produce ends corresponding to the sticky ends of the SC adaptors; two class IIS restriction endonucleases that recognize restriction sites on the SC adaptor and the amplification adaptor; T4 DNA ligase; and an agent for polymerization of the nucleoside triphosphates. Instructions providing information to the user regarding the use of the kit can also be inserted, including for identifying polymorphic regions in bacterial genomic DNA. The kit may also include biotinylated amplification and adaptor primers; and streptavidin beads for the purification of biotinylated products.

The following specific examples are provided to further illustrate the invention.

One skilled in the art can readily appreciate that various modifications can be made to the described embodiments without departing from the scope and spirit of the invention. Such modifications are also intended to be within the scope of the invention.

Examples

Example 1 demonstrates the sequence capturing function of the SC adaptor and the use of A-T-P as extension primer. Example 2 illustrates the functional reversal of A-T-P by ligation to the reverser adaptor and PCR amplification. Example 3 shows the creation of *in vitro* bacterial libraries and the preparation of strain-specific A-T-Ps. It also demonstrates the utility of A-T-Ps in the identification of polymorphic regions. Example 4 illustrates the use of A-T-Ps in bacterial strain typing.

Example 1. Sequence Capturing function of the SC adaptor and the use of A-T-P as extension primer

1. Preparation of the SC adaptor and PCR primers

The SC adaptor used in our experiments is shown in Figure 1 (SEQ ID NO: 1). It is a single long oligonucleotide that can fold itself into a partially double-stranded structure. GC-rich regions are located at each end of the adaptor. Between these GC-rich regions, a "bubble" of non-complementary sequence is located in the center. The sequence on the upper region is termed the SS2 sequence, and the lower region is termed the SS8 region. Adaptor-specific PCR primers with the same sequences as the SS2 and SS8 are called PSS2 and PSS8, respectively (Table 1). Note that the sequences of the PCR primers are always in the same 5' to 3' direction as the non-complementary sequence of the SC adaptor. Restriction endonuclease recognition sites are present in the SC adaptor. Kpn2I and Eco31I were incorporated for diagnostic purposes. SalI, Cfr42I, Eco52I and NotI are present for cloning the PCR products if required. GsuI is the class IIS restriction site for capturing sequence adjacent to the adaptor. The open end of the adaptor has a sticky end that is complementary to the sticky end of a particular restriction endonuclease of interest. When ligated to a restriction digested DNA end, the restriction endonuclease recognition site is reconstituted. Four SC adaptors with sticky ends corresponding to PstI, EcoRI, HindIII and PagI were used in our experiments for creating the *in vitro* libraries.

The SC adaptor oligonucleotide and PCR primers were chemically synthesized by MOBIX (McMaster University, Hamilton, Canada). Before using the SC adaptor for ligation to DNA fragments, the oligonucleotide was heated at 85°C for 5 minutes at a concentration of 25 pmol/μl. The oligonucleotide was then allowed to cool slowly down to room temperature. These denaturation/renaturation steps allowed the oligonucleotide to fold into the desired "bubble" shape. The PCR primers were diluted to a working concentration of 25 pmol/μl in ddH_2O. The sequences of the adaptors and the primers are listed in Table 1.

Table 1. Sequences of the adaptors and PCR primers used in Examples 1-4

Adaptor/Primer Name	DNA Sequence
PstI SC Adaptor (5' phosphorylated)	5'- GCTCCAGCGGCCGCGGTCGACGTACGATCAGACGCATAGG TCTCCGG ATTTTCCGGAGACCTAGATAGTCTGCGTCGACAGTCCCGC GGCCGCTGGAGCTGCA -3' (SEQ ID NO: 2)

EcoRI SC Adaptor (5' phosphorylated)	5'- AATTCCTCCAGCGGCCGCGGTCGACGTACGATCAGACGCA TAGGTCT CCGGATTTTCCGGAGACCTAGATAGTCTGCGTCGACAGTC CCGCGGCCGCTGGAGG -3' (SEQ ID NO: 3)
HindIII SC Adaptor (5' phosphorylated)	5'- AGCTTCTCCAGCGGCCGCGGTCGACGTACGATCAGACGCA TAGGTCT CCGGATTTTCCGGAGACCTAGATAGTCTGCGTCGACAGTC CCGCGGCCGCTGGAGA -3' (SEQ ID NO: 4)
PagI SC Adaptor (5' phosphorylated)	5'- CATGACTCCAGCGGCCGCGGTCGACGTACGATCAGACGCA TAGGTCT CCGGATTTTCCGGAGACCTAGATAGTCTGCGTCGACAGTC CCGCGGCCGCTGGAGT -3' (SEQ ID NO: 5)
PSS2	5'- GATAGTCTGCGTCGACAGTC -3' (SEQ ID NO: 6)
PSS4	5'- CAACGTGAGCGATGTCTAGT -3' (SEQ ID NO: 7)
3'DNAK	5'- TGCAGTCAGCGCAGACTC -3' (SEQ ID NO: 8)
5'DNAK	5'- GCCGCAGATCGAAGTTAC -3' (SEQ ID NO: 9)
5'DNAK2	5'- GATTATGGATGGCACCACTC -3' (SEQ ID NO: 10)
3'IS431	5'- GTTGAAGGTGCCTGATCTGT -3' (SEQ ID NO: 11)
3'DEGIS431	5'- TGGTGCATATTCTTGAACCCANCKRTA -3' N = A, G, C, or T; K = G or A; R = G or A (SEQ ID NO: 12)
5'IS431	5'- GTAGCCGTTGGCTACTATCTAAG -3' (SEQ ID NO: 13)

2. Restriction endonuclease digestion of bacterial genomic DNA

E. coli genomic DNA was isolated from 10 ml of overnight DH5α culture following the manufacturer's protocol (Qiagen). The final concentration of the purified genomic DNA was 655 μg/ml. An aliquot of the purified genomic DNA (~10 ng) was digested with 50 U of Pst I in 1X O buffer (Fermentas) at 37°C for 1 hour. The digested genomic DNA was purified with phenol/chloroform mixture and ethanol precipitated. The DNA pellet was resuspended in 20 μl of ddH_2O. The final DNA concentration was estimated to be 288 fmol ends/μl.

Table 2. PstI digestion of E. coli genomic DNA

Components	Volume (µl)	Final Amount
Bacterial Genomic DNA (655 ng/µl)	15	9.8 µg
10X O^+ Buffer (500mM Tris-HCl (pH7.5), 100mM $MgCl_2$, 1000mM NaCl and 1mg/ml BSA)	5	1X
Pst I (10 U/µl)	5	50 U
ddH_2O	25	
	50	

3. Ligation of the SC adaptor to restriction digested DNA fragments

The Pst I-digested genomic DNA (1 pmol ends) was mixed with pre-treated SC adaptor (3 pmol) in 1X T4 DNA ligase reaction buffer (Fermentas). Ten units of T4 DNA ligase were added and the ligation reaction was carried out overnight at room temperature. The ligation reaction was heated at 65°C for 10 minutes to inactivate the T4 ligase.

Table 3. Ligation of PstI-digested genomic DNA to SC adaptor

Components	Volume (µl)	Final Amount
Pst I digested DNA (288 fmol ends/µl)	3.5	1 pmol ends
SC Adaptor-Pst I (2.5 pmol/µl)	1.2	3 pmol
10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	2.5	1X
T4 DNA ligase (5U/µl)	2	10 U
ddH_2O	15.8	
	25	

4. PCR amplification of unique fragment containing the SC adaptor sequence

To demonstrate that the SC adaptor can capture a unique sequence and be used as extension primer, a 20 bp oligonucleotide was synthesized corresponding to the E. coli dnaK gene as documented in the GenBank. The primer sequence (3'DNAK) is shown in Table 1. This primer was used in a PCR reaction with the SC adaptor primer (PSS2) to amplify a portion of the dnaK gene from the *in vitro* Pst I genomic library created in step 3 above. The amplified DNA fragment would correspond to the part of the dnaK gene from the closest Pst I site to the location of the DnaK primer sequence (See Figure 6). The expected size of the amplified product should be 375 bp long. As a positive control primer,

another oligonucleotide (5'DNAK) corresponding to a 20 bp sequence in front of the 3'DNAK primer was also synthesized (Table 1). When the 5' DNAK and 3'DNAK primers were used in a PCR reaction to amplify the portion of the dnaK gene from the *in vitro* Pst I genomic library, an amplified product of 290 bp long would be obtained.

The PCR reaction was set up as shown in the Table 4 below. The PCR amplification was carried out in a Perkin Elmer thermocycler under the following conditions: 94°C for 3 minutes; 30 cycles of 94°C for 30 seconds, 65°C for 15 seconds and 72°C for 1 minute; 72°C for 5 minutes. After PCR, 10 µl of the reaction was ran on a 1% ethidium bromide pre-stained agarose gel at120 V for 1 hour. A photograph of the gel was taken under the illumination of an ultraviolet light.

For both the sample reaction and the positive control reaction, only a single DNA band was observed for each reaction. With the help of the DNA molecular weight markers, these bands were estimated to be ~ 370 bp and ~ 290 bp for the sample reaction and the positive control reaction respectively. Therefore, the portion of the dnaK gene between the PstI restriction site and the 3'DNAK primer sequence was successfully amplified.

Table 4. PCR amplification of *E. coli in vitro* PstI library

	Sample RXN		Positive Control RXN	
Components	Volume (µl)	Final Conc.	Volume (µl)	Final Conc.
In Vitro PstI genomic Library (50 ng/µl)	1	1 ng/µl	1	1 ng/µl
PSS2 primer (25 pmol/µl)	1	0.5 pmol/µl	-	-
3' DNAK primer (25 pmol/µl)	1	0.5 pmol/µl	1	0.5 pmol/µl
5' DNAK primer (25 pmol/µl)	-	-	1	0.5 pmol/µl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	5	1X	5	1X
$MgCl_2$ (25 mM)	3	1.5 mM	3	1.5 mM
dNTPs (2mM each)	5	200 µM	5	200 µM
Taq DNA Polymerase (5 U/µl)	1	0.1 U/µl	1	0.1 U/µl
ddH_2O	33		33	
	50		50	

5. A-T-P preparation

a) Purification and digestion of the amplified product

The sample PCR reaction (40 µl) was run on a 1% agarose gel as described before. The amplified product was cut out from the gel and purified with the DNA Extraction kit (Fermentas). The purified product was estimated to have a concentration of 10 pmol/µl. The purified DNA fragment was digested with Gsu I in 1X B^+ buffer (Fermentas) at 30°C overnight to release the SC adaptor with the "captured sequence".

Table 5. Class IIS RE digestion of the amplified product from *in vitro* library

Components	Volume (µl)	Final Amount
PCR Fragment (10 pmol/µl)	20	200 pmol
10X B Buffer (100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$ and 1mg/ml BSA)	6	1X
Gsu I (1 U/µl)	5	5 U
ddH_2O	29	
	60	

The digestion reaction was inactivated at 65°C for 20 minutes. Double-distilled water was added to make the final volume to 100 µl. The diluted reaction was spun through a Nanosep (Gelman) 100K MW cut-off filter unit at 5000 rpm in a microcentrifuge for 20 minutes. This step served to separate most of the long PCR product and the Gsu I restriction endonuclease from the A-T-P_{DNAK} (the SC adaptor with the captured sequence). The filtrate containing the A-T-P_{DNAK} was recovered. The final concentration of the SC adaptor was estimated to be 2 fmol/µl.

(b) Ligation of the A-T-P to the amplification adaptor

The A-T-P_{DNAK} was ligated to the amplification adaptor (SEQ ID NOS: 14 and 15) for 2 hours at room temperature in a reaction as described in Table 6 below. The ligation reaction was heated at 65°C for 10 minutes to inactivate the ligase.

Table 6. Ligation of the A-T-P to the amplification adaptor

Components	Volume (µl)	Final Amount
A-T-P_{DNAK} (2 fmol /µl)	20	40 fmol
Amplification Adaptor (1 pmol/µl)	1.92	1920 fmol

10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	3	1X
T4 DNA ligase (5U/μl)	2	10 U
ddH_2O	3.08	
	30	

(c) Reproduction of the A-T-P by PCR amplification

By ligating the amplification adaptor to the A-T-P_{DNAK}, the sequences at both ends of the ligation product were now known. By using PCR primers complementary to these ends, the ligation product may be reproduced by PCR. PCR was carried out using the SC adaptor primer, PSS2 and the amplification adpator primer, PSS4, in a reaction as described in the Table 7. The PCR conditions was as follows: 94°C for 30 seconds; 10 cycles of 94°C for 5 seconds, 58°C for 5 seconds, and 72°C for 10 seconds; 72°C for 1 minute.

Table 7. PCR amplification of the A-T-P/amplification adaptor complex to prepare A-T-P stock

Components	Volume (μl)	Final Concentration
A-T-P_{DNAK} Ligation (1.33 fmol/μl)	1.5	0.04 fmol/μl
PSS2 primer (25 pmol/μl)	1	0.5 pmol/μl
PSS4 primer (25 pmol/μl)	1	0.5 pmol/μl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	5	1X
$MgCl_2$ (25 mM)	3	1.5 mM
dNTPs (2mM each)	5	200 μM
Taq DNA Polymerase (5 U/μl)	1	0.1 U/μl
ddH_2O	32.5	
	50	

After PCR, this reaction served as the stock solution for A-T-P_{DNAK}. To prepare A-T-P_{DNAK} for extension/PCR reactions, a second round of PCR amplification was carried out with an aliquot of the first-round PCR reaction (Table 8). In the second round PCR, the same PCR primer pair was used except that the PSS4 primer was biotinylated. The same PCR conditions as the first-round PCR were used except the total number of cycle was increased to 20. A 10 μl aliquot of the second-round PCR reaction was ran on a 20% polyacrylamide gel in TBE buffer at120V for 90 minutes. After electrophoresis, the gel was stained with ethidium bromide and then a picture of the gel was taken. The 76 bp amplified ligation complex of A-T-P_{DNAK} and the amplification adaptor was observed.

Table 8. PCR amplification to prepare working solution of A-T-P

Components	Volume (μl)	Final Concentration
First-round PCR reaction	1	
PSS2 primer (25 pmol/μl)	2	0.5 pmol/μl
Biotinylated PSS4 primer (25 pmol/μl)	2	0.5 pmol/μl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	10	1X
$MgCl_2$ (25 mM)	6	1.5 mM
dNTPs (2mM each)	10	200 μM
Taq DNA Polymerase (5 U/μl)	2	0.1 U/μl
ddH_2O	67	
	100	

To further remove any long contaminating DNA fragments, second-round PCR reactions (200 μl) was ran on a 2.5% agarose gel. The biotinylated PCR product of 76 bp was cut from the gel. The agarose slice containing the PCR product was spun in a spin-filter unit at 1000xg for 10 minutes. The biotinylated PCR product went through the filter with the liquid and was collected. The biotinylated PCR product was ethanol precipitated and resuspended in 100 μl of ddH_2O. The concentration of the biotinylated PCR product was estimated to be about 80 fmol/μl.

(d) Binding of the A-T-P/Amplification adaptor complex to Streptavidin beads

Approximately 50 μl of Streptavidin beads (Dynal) were washed 2 times with 150 μl of 6X SSC buffer (0.9M NaCl, 0.09M Sodium citrate, pH 7.0), and then 2 times with 150 μl of 2X B&W buffer (10mM Tris-HCl (pH7.5), 1mM EDTA and 2M NaCl). After removing the wash liquid, the beads were resuspended in 50 μl of 2X B&W buffer. Fifty μl of the biotinylated PCR product was added to the Streptavidin beads in a microcentrifuge tube. The tube was taped to a rotating stage for 30 minutes at room temperature to ensure constant mixing. After the incubation period, the beads were washed 3 times with 150 μl of 2X B&W buffer, and then 3 times with 100 μl of 1X B buffer. The wash buffer was completely removed from the tube after the last wash.

(e) Release of the A-T-P from the bead-bound A-T-P/Amplification adaptor complex

The A-T-P/Amplification adaptor complex bound beads were treated with Gsu I to release the A-T-P$_{DNAK}$. The GsuI digestion was set up as shown in the Table 9. The reaction was carried out at 30°C for 2 hours. The Gsu I enzyme was inactivated by heating at 65°C for 20 minutes. Chloroform may be used to purify and concentrate the A-T-P$_{DNAK}$. The final concentration of the A-T-P$_{DNAK}$ was estimated to be about 30-50 fmol/µl.

Table 9. Release of functional A-T-P from Streptavidin beads

Components	Volume (µl)	Final Amount
A-T-P/Adaptor complex bound beads		200 pmol
10X B^+ Buffer (100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$ and 1mg/ml BSA)	8	1X
Gsu I (1 U/µl)	6	6 U
ddH_2O	66	
	80	

6. Extension/PCR of A-T-P on undigested bacterial genomic DNA

To demonstrate that the A-T-P$_{DNAK}$ was functional as an extension primer, the purified A-T-P$_{DNAK}$ was used to amplify a part of the dnak gene from E. coli genomic DNA in a novel extension/PCR reaction. The first part of the reaction was the primer extension of the A-T-P$_{DNAK}$ on the genomic DNA using Taq DNA polymerase. The captured dnak gene sequence of the A-T-P$_{DNAK}$ would find its complementary sequence on the genomic DNA and hybridize to it. Taq DNA polymerase then extended the A-T-P$_{DNAK}$ primer by copying the complementary sequence. The newly synthesized strand would contain the complementary sequence of the 3'DNAK primer. The second part of the reaction was the PCR amplification of the portion of the dnak gene, defined by the A-T-P$_{DNAK}$ priming site (i.e. the Pst I restriction site) and the 3'DNAK primer binding site, using the PSS2 and 3'DNAK primers.

Fifty ng of purified E. coli genomic DNA was mixed with about 40 fmol of the A-T-P$_{DNAK}$ in a 1X PCR reaction buffer. Five units of Taq DNA polymerase were used for the reaction. As a negative control reaction, another reaction mixture was also set up but without the genomic DNA as template. The reactions (Table 10) were repeated 5 times under the following temperature profiles in a thermocycler: One minute at 94°C, 1 minute at 37°C and 1 minute at 72°C. After this extension step, the PSS2 primer and the 3'DNAK

primer were added and PCR (Table 11) was carried out under the following conditions: 94°C for 3 minutes; 30 cycles of 94°C for 30 seconds, 62°C for 1 minute, and 72°C for 2 minutes; 72 minutes for 5 minutes.

Table 10. Extension of A-T-P on genomic DNA

Components	Sample RXN		Negative Control RXN	
	Volume (µl)	Final Conc.	Volume (µl)	Final Conc.
A-T-P_{DNAK} (30-50 fmol/µl)	1.7	2.5-4.25 fmol/µl	1.7	2.5-4.25 fmol/µl
E. coli genomic DNA (50 ng/µl)	1	2.5 ng/µl	-	-
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	2	1X	2	1X
$MgCl_2$ (25 mM)	1.2	1.5 mM	1.2	1.5 mM
dNTPs (2mM each)	2	200 µM	2	200 µM
Taq DNA Polymerase (5 U/µl)	1	0.1 U/µl	1	0.1 U/µl
ddH_2O	11.1		12.1	
	20		20	

Table 11. PCR amplification of the extension reaction

Components	Volume (µl)	Final Concentration
Extension reaction	20	
PSS2 (25 pmol/µl)	1	0.5 pmol/µl
3'DNAK primer (25 pmol/µl)	1	0.5 pmol/µl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	3	1X
$MgCl_2$ (25 mM)	1.8	1.5 mM
dNTPs (2mM each)	3	200 µM
ddH_2O	20.2	
	50	

After PCR, 10 µl aliquots of the reactions were run in a 1% agarose gel. Only the reaction with the E. coli genomic DNA had an amplified DNA band of ~370 bp. This DNA band was cut from the gel and purified. The purified DNA was sent for sequencing. The sequence of this band matched perfectly with the expected DNA sequence of the dnak gene as documented in GenBank.

This example demonstrated that the SC adaptor was capable of capturing sequence from the DNA fragments to which it was ligated. The SC adaptor with the captured sequence (A-T-P) was functional as sequence-specific primer in a novel extension/PCR reaction. This example also demonstrated that the A-T-P could be reproduced using an amplification adaptor. In the present invention, the sequence of the captured sequence is not required to be known before the extension/PCR step.

Example 2. Functional reversal of A-T-P

At times, it may be useful to be able to use the A-T-P to prime DNA synthesis in the opposite direction. For example, this may allow capture of further upstream sequences. The functional direction of the A-T-P needs to be reversed in order to do so. This can be accomplished with a slight modification in Step 5 (c) & (e) of Example 1.

To prepare "reversed" A-T-P, biotinylated PSS2 primer was used instead of PSS4 for the second-round PCR in step 5(c) of Example 1. This put the biotin moiety at the A-T-P end of the A-T-P/amplification adaptor complex. After binding the complex to streptavidin beads, the complex was digested with the Type IIS restriction endonuclease which recognize the restriction site on the amplification adaptor. The captured sequence was released together with the amplification adpator from the beads. This complex is termed the "reversed" A-T-P. It is functional in primer extension in the opposite direction as compared to the original A-T-P.

1. Amplification of A-T-P_{DNAK}/Amplification Adaptor complex

The stock A-T-P_{DNAK} solution in step 5(c) of Example 1 was used in a second round PCR with biotinylated PSS2 primer as described in Table 12. The PCR conditions were exactly the same as in the second round PCR of step 5(c) in Example 1.

Table 12. PCR amplification of A-T-P to prepare "reversed" A-T-P

Components	Volume (μl)	Final Concentration
First-round PCR reaction	1	
PSS4 primer (25 pmol/μl)	2	0.5 pmol/μl
Biotinylated PSS2 primer (25 pmol/μl)	2	0.5 pmol/μl

10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	10	1X
$MgCl_2$ (25 mM)	6	1.5 mM
dNTPs (2mM each)	10	200 µM
Taq DNA Polymerase (5 U/µl)	2	0.1 U/µl
ddH_2O	67	
	100	

2. Purification and Binding of the A-T-P/amplification adaptor complex

After PCR, the reaction was run in a 2.5% agarose gel to purify the amplified product as described in Step 5(c) of Example 1. The purified A-T-P/amplification adaptor complex was mixed with streptavidin beads and binding was carried out as described in step 5(d) of Example 1.

3. Release of the reversed A-T-P_{DNAK}

The beads with bound A-T-P/amplification adaptor complex were digested with Eco 57I that recognized the restriction site on the amplification adaptor (Table 13). After digestion at 37°C for 2 hours, the captured sequence was released together with the amplification adaptor. This molecule may now be used as "reversed" A-T-P.

Table 13. Release of "reversed" A-T-P from Streptavidin beads

Components	Volume (µl)	Final Amount
A-T-P/Adaptor complex bound beads		200 pmol
10X B Buffer (100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$ and 1mg/ml BSA)	8	1X
50X SAM (0.5 mM)	1.6	0.01 mM
Eco57 I (5 U/µl)	0.8	4 U
ddH_2O	69.6	
	80	

4. Extension/PCR of A-T-P on undigested bacterial genomic DNA

To demonstrate that the reversed A-T-P was functional in primer extension, it was used in the extension of a portion of the dank gene in the opposite direction compared to the extension direction in step 6 of Example 1 (Figure 6). A new PCR primer that was located 5' to the priming site of the reversed A-T-P (i.e. the Pst I site) on the dnaK gene was synthesized (Table 1). This PCR primer (5'DNAK2) and the amplification primer (PSS4)

would be able to amplify a DNA fragment of ~1270 bp long if an extension product of reversed A-T-P was present after primer extension.

Primer extension was carried out with the reversed A-T-P_{DNAK} on E. coli genomic DNA as described in the Table 14. The reaction conditions were the same as in step 6 of Example 1.

Table 14. Extension of "reversed" A-T-P on genomic DNA

	Sample RXN		Negative Control RXN	
Components	Volume (µl)	Final Conc.	Volume (µl)	Final Conc.
reversed A-T-P_{DNAK} (30-50 fmol/µl)	1	2.5-4.25 fmol/µl	1.7	2.5-4.25 fmol/µl
E. coli genomic DNA (50 ng/µl)	1	2.5 ng/µl	-	-
10X PCR Buffer (100mM Tris-HCl (pH 7.5), 100mM $MgCl_2$ and 1mg/ml BSA)	2	1X	2	1X
$MgCl_2$ (25 mM)	1.2	1.5 mM	1.2	1.5 mM
dNTPs (2mM each)	2	200 µM	2	200 µM
Taq DNA Polymerase (5 U/µl)	1	0.1 U/µl	1	0.1 U/µl
ddH_2O	11.8		12.1	
	20		20	

After primer extension, the PSS4 primer and the 5'DNAK2 primer were added to the reaction mixture (Table 15). PCR was carried out the same way as described in step 6 of Example 1.

Table 15. PCR amplification of the extension reaction

Components	Volume (µl)	Final Concentration
Extension reaction	20	
PSS4 (25 pmol/µl)	1	0.5 pmol/µl
5'DNAK2 primer (25 pmol/µl)	1	0.5 pmol/µl
10X PCR Buffer (	3	1X
$MgCl_2$ (25 mM)	1.8	1.5 mM
dNTPs (2mM each)	3	200 µM
ddH_2O	20.2	
	50	

After PCR, 10 μl aliquots of the reactions were run on a 1% agarose gel. Only the sample reaction had an amplified product of ~ 1270 bp long. This result confirmed that the reversed A-T-P_{DNAK} was functional in primer extension. It demonstrated that A-T-P may be functional reversed as described.

Example 3. Creation of *in vitro* bacterial genomic libraries and preparation of strain-specific A-T-Ps for S. aureus

Staphylococcus aureus is one of the major causes of noscomial infections and outbreaks in hospitals. Due to the epidemiological significance of *S. aureus*, many typing methods have been developed and standardized using this organism. Therefore, S. aureus was used as a model in this example.

1. Restriction endonuclease digestion of bacterial genomic DNA

Four strains of *S. aureus*, isolated during independent outbreaks at NIH hospitals, were obtained. The four strains were all tested to be methicillin resistance at NIH. Genomic DNAs were purified from 10 ml of overnight cultures of these bacterial strains. Pelleted cells were washed and resuspended in 1 ml of TE buffer (pH 7.5). A 200 μl aliquot of each cell suspensions was mixed with 20 μl of Lysostaphin (0.01 μg/μl) and 50 μl of lysozyme (50 mg/ml). The mixtures were incubated at 37°C for 1 hour to break up the cells. After the incubation period, 30 μl of 10% SDS and 50 μl of 7.5 M Sodium acetate were added to each mixture. The mixtures were phenol/chloroform extracted. The purified genomic DNAs were ethanol precipitated and resuspended in 100 μl. The final concentrations of the genomic DNA from the 4 strains were between 0.4 - 1 μg/μl.

Complete digestions of the genomic DNAs were individually carried out with HindIII, EcoRI, PagI, and PstI in separate reactions as described in the Tables 16-19 below. The restriction digestions were carried out at 37°C for 2 hours. After digestion, the reactions were inactivated at 65°C for 10 minutes. The restriction digested DNAs were purified with the DNA extraction kit (Fermentas) and resuspended in 40 μl of ddH_2O.

Table 16. Digestion of *S. aureus* DNA with Hind III

Components	Volume (µl)	Final Amount
S. aureus Genomic DNA (500 ng/µl)	5	2.5 µg
10X R^+ Buffer (100mM Tris-HCl (pH8.5), 100mM $MgCl_2$, 1000mM KCl and 1 mg/ml BSA)	2.5	1X
Hind III (10 U/µl)	1	10 U
ddH_2O	16.5	
	25	

Table 17. Digestion of *S. aureus* DNA with Eco RI

Components	Volume (µl)	Final Amount
S. aureus Genomic DNA (500 ng/µl)	5	2.5 µg
10X O^+ RI Buffer (500mM Tris-HCl (pH7.5), 100mM $MgCl_2$, 1000mM NaCl and 1mg/ml BSA)	2.5	1X
Eco RI I (10 U/µl)	1	10 U
ddH_2O	16.5	
	25	

Table 18. Digestion of *S. aureus* DNA with Pag I

Components	Volume (µl)	Final Amount
S. aureus Genomic DNA (500 ng/µl)	5	2.5 µg
10X O Buffer (500mM Tris-HCl (pH7.5), 100mM $MgCl_2$, 1000mM NaCl and 1mg/ml BSA)	2.5	1X
Pag I (10 U/µl)	1	10 U
ddH_2O	16.5	
	25	

Table 19. Digestion of *S. aureus* DNA with Pst I

Components	Volume (µl)	Final Amount
S. aureus Genomic DNA (500 ng/µl)	5	2.5 µg
10X O Buffer (500mM Tris-HCl (pH7.5), 100mM $MgCl_2$, 1000mM NaCl and 1mg/ml BSA)	2.5	1X
Pst I (10 U/µl)	0.5	5 U
ddH_2O	17	
	25	

2. Ligation of SC adaptor to RE-digested genomic DNA

The digested genomic DNAs were ligated to SC adaptors with sticky ends corresponding to the respective cuts. Based on the cutting frequency of each enzyme on the S. aureus genome, the number of ends in pmol was estimated for each digested genomic

DNA. The ligation reactions were set up at a molar ratio of 3:1 (adaptor ends: DNA fragment ends) as described in Tables 20-23 below. The ligation reactions were carried out at room temperature overnight. The reactions were then inactivated at 65°C for 10 minutes. The ligation experiments resulted in 4 different *in vitro* libraries for each of the 4 strains.

5

Table 20. Ligation of Hind III-SC adaptor to Hind III-digested genomic DNA

Components	Volume (µl)	Final Amount
S. aureus -HindIII digested DNA (40 ng.µl)	25	1 µg
SC Adaptor-HindIII (1 pmol/µl)	3.6	3.6 pmol
10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	4	1X
T4 DNA ligase (5U/µl)	4	20 U
ddH_2O	3.4	
	40	

Table 21. Ligation of Eco RI-SC adaptor to Eco RI-digested genomic DNA

Components	Volume (µl)	Final Amount
S. aureus -EcoRI digested DNA (40 ng.µl)	20	1 µg
SC Adaptor-Eco RI (1 pmol/µl)	2	2 pmol
10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	4	1X
T4 DNA ligase (5U/µl)	4	10 U
ddH_2O	5	
	40	

Table 22. Ligation of Pag I-SC adaptor to Pag I-digested genomic DNA

Components	Volume (µl)	Final Amount
S. aureus -Pag I digested DNA (40 ng.µl)	25	1 µg
SC Adaptor-Pag I (1 pmol/µl)	2.9	2.9 pmol
10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	4	1X
T4 DNA ligase (5U/µl)	4	10 U
ddH_2O	4.1	
	40	

Table 23. Ligation of Pst I-SC adaptor to Pst I-digested genomic DNA

Components	Volume (µl)	Final Amount
S. aureus -Pst I digested DNA (40 ng.µl)	25	1 µg
SC Adaptor-Pst I (1 pmol/µl)	1.2	1.2 pmol
10X T4 DNA ligase Buffer (400mM Tris-HCl (pH7.8), 100mM $MgCl_2$, 100mM DTT and 5mM ATP)	4	1X
T4 DNA ligase (5U/µl)	4	10 U

ddH_2O	5.8 40	

3. PCR amplification of unique fragments from *in vitro* bacterial genomic libraries

Because of the vast amount of data available on IS elements, IS element was chosen as the anchor sequence in our search for polymorphic regions in *S. aureus*. IS257 element was found in direct repeats flanking the trimethoprim-resistance determinant on many members of the pSK1 family of multi-resistance plasmids in *S. aureus*. It also forms direct repeats bounding mercury-resistance determinants on heavy-metal resistance plasmids; these repeats are also known as IS431 (which is an isoform of IS257 in the IS6 family). In addition, IS257 sequences are found adjacent to chromosomal determinants for resistance to mercury, methicillin and tetracycline. Since the bacteria strains obtained from NIH are all methicillin resistance, IS257 seemed like a good candidate.

A PCR primer with sequence complementary to an internal region of the transposase gene in IS431 insertion element was prepared. The priming direction of the primer was in the opposite direction of the open reading frame of the insertion element. The sequence information of the insertion sequence was obtained from the IS431mec gene associated with methicillin resistance in *S. aureus* (GenBank accession # X53818). The anchor primer (3'IS431) sequence is shown in Table 1. A 3' degenerate IS431 primer (3'DEGIS431) based on a conserved region of the IS431/IS257 transposase family was also prepared. The degenerate primer was designed according to the COnsensus-DEgenerate Hybrid Oligonucleotide Primers (CODEHOP) design strategy described in Rose, et al. Nucleic Acids Res. 26:1628-1635 (1998). A 5' primer (5' IS431) was also designed which is primarily used to provide a positive test for the presence of IS431 transposase sequence.

The uncut genomic DNAs were tested for the presence of the IS431 element using the 5'IS431 and 3'IS431 primers in PCR reactions. The corresponding 368 bp fragment of the transposases gene was obtained for all of the four strains tested.

The ligation reactions described above may be used without the need for further purification. The ligation products were amplified with the adaptor primer and the anchor

primer (3'IS431) in PCR reactions as described in Table 24. As a negative control, S. aureus undigested and unligated genome DNA was used in place of the *in vitro* library in one of the PCR reaction. As a positive PCR control, 5'IS431 and 3'IS431 primers were used to amplify a fragment of the IS431 transposase gene from S. aureus genomic DNA. The PCR reaction conditions were: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds, and 72°C for 5 minutes; and 72°C for 5 minutes. Ten µl aliquots of each reaction were run on a 1% agarose gel.

Table 24. PCR amplification of *S. aureus in vitro* libraries

	Volume (µl)						Final Conc.
Components	1	2	3	4	5	6	
S. aureus strain 1-HIndIII *in vitro* library (25 ng/µl)	1	-	-	-	-	-	0.5 ng/µl
S. aureus strain 1-EcoRI *in vitro* library (25 ng/µl)	-	1	-	-	-	-	0.5 ng/µl
S. aureus strain 1-Pag I *in vitro* library (25 ng/µl)	-	-	1	-	-	-	0.5 ng/µl
S. aureus strain 1-Pst I *in vitro* library (25 ng/µl)	-	-	-	1	-	-	0.5 ng/µl
S. aureus genomic DNA (25 ng/µl)	-	-	-	-	1	1	0.5 ng/µl
PSS2 Primer (25 pmol/µl)	1	1	1	1	1	-	0.5 pmol/µl
3'IS431 Primer (25 pmol/µl)	1	1	1	1	1	1	0.5 pmol/µl
5'IS431 Primer (25 pmol/µl)	-	-	-	-	-	1	0.5 pmol/µl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	5	5	5	5	5	5	1X
$MgCl_2$ (25 mM)	3	3	3	3	3	3	1.5 mM
dNTPs (2mM each)	5	5	5	5	5	5	200 µM
Taq DNA Polymerase (5 U/µl)	1	1	1	1	1	1	0.1 U/µl
ddH_2O	33	33	33	33	33	33	
	50	50	50	50	50	50	

All of the four reactions with *in vitro* libraries of strain 1 produced amplified products. In similar experiments using *in vitro* libraries from the other 3 strains, amplified products were also obtained for the *in vitro* libraries of these strains. A degenerate 3'IS431 primer was also used as the anchor primer (in place of 3'IS431) in the PCR reactions. The

summary of the amplification results for all of the *in vitro* libraries of the four strains were shown in Table 25.

As can be seen from Table 25, amplified products of different lengths were obtained from different libraries for each strain. The lengths of the amplification products for a particular library were sometimes the same between two strains but never the same for all four strains. The exception was the HindIII libraries. This was due to the presence of a HindIII site just 200 bp in front of the 3'IS431 primer. When 3'DEGIS431 (DE) primer was used as the anchor primer in the PCR reactions of the HindIII libraries, the lengths of the amplified products were again different among the strains; Strain 1 and Strain 4 shared a common 1900 bp band, and Strain 2 and Strain 3 shared a common 800 bp band. The right side of the Table 24 shows the results obtained when the 3' degenerate primer was used as the anchor primer. From documented sequence data, the positions of the SS primer and the DE primer are about 250 bp apart. In all cases, the amplified bands obtained from PCR reactions using the DE primer were judged to be fairly similar in size to those obtained from the SS primer and we adopted a general practice of labelling the length of the amplified products obtained from DE primer as 250 bp smaller than the corresponding bands obtained from SS primer.

As a first step in determining if the amplified products obtained for a particular library represent a variable region, we sequenced the amplified products obtained from the PstI library for all 4 strains (using the SS anchor primer). The sequence of the 2000 bp product from Strain 1 was found to be exactly the same as that of Strain 4. Likewise, the sequence of 900 bp product of Strain 2 was found to be exactly the same as that of Strain 3. In addition, the 5' and 3' ends of

Table 25. PCR Products from *in vitro S. aureus* DNA libraries using Adaptor primer and Anchor primer (SS or DE)

Library	3'IS431 (SS)[1]				3'DEGIS431 (DE)[1]			
	Strain 1	Strain 2	Strain 3	Strain 4	Strain 1	Strain 2	Strain 3	Strain 4
Pst I	2000 bp	900 bp	900 bp	2000 bp	1750 bp	650 bp	650 bp	1750 bp
Eco RI	4000 bp	1900 bp	1900 bp 1400 bp	1400 bp	3750 bp	1650 bp	1650 bp 1150 bp	1150 bp
Pag I	1900 bp 800 bp[3]	2000 bp 800 bp	800 bp[3]	1900 bp	1650 bp 550 bp[3]	1750 bp 550 bp	550 bp[3]	1650 bp
Hind III[2]	200 bp	200 bp	200 bp	200 bp	2500 bp 1900 bp	800 bp	800 bp 600 bp	1900 bp

[1] The sizes of the PCR products for 3'IS431 reactions were estimated from MW markers ran on agarose gels; for 3'DEGIS431 reactions, the sizes of the PCR products were estimated and then corrected to match the difference of 250 bp between the SS and DE primers

[2] There is a HindIII site, about 200 bp, in front of the SS anchor primer; DE anchor primer is about 250 bp in front of the SS anchor primer

[3] These bands are composed of multiple fragments of similar sizes that can be resolved on higher percentage agarose gels

all of these 4 products were the same. The 1100 bp difference between the two different product lengths was due to the different insertion sites of the IS431 element in different strains. For Strain 2 and 3, the IS431 inserted at a location that is 1100 bp upstream of the insertion site of IS431 for Strain 1 and Strain 4. The 3' end sequence homology for all of the strains represent the common sequence (~470 bp) of the IS431 insertion element.

We then compared the sequences we obtained with all known DNA sequences documented in the GenBank by using the BLAST program at the NCBI web site. We found that the sequences matched the mec DNA region of methicillin-resistant *Staphylococcus aureus* (GenBank Accession #D86934). This is the region of DNA that is responsible for methicillin resistant and is known to be associated with IS431 elements. This confirmed the specific amplification achieved by the present method.

4. Preparation of strain-specific A-T-Ps

While the initial sequencing was carried out to confirm the integrity of the methodology, the invention contemplates the preparation of A-T-Ps without sequencing.

The next step in our experiments was to prepare A-T-P from the amplified products obtained above. The preparation of A-T-P was carried out as described in step 5 of Example 1 and shown in Figure 5. Thirteen A-T-Ps were prepared from the amplification products of the *in vitro* libraries of the 4 S. aureus strains.

Only one of the 13 A-T-P did not produced amplified product (i.e. not functional) in extension/PCR reactions. The cause of this was probably due to the captured sequence. The captured sequence for this particular A-T-P has a stable internal dimer structure that would interfere with the priming during extension/PCR reactions. Two of the 13 A-T-Ps only produced faint bands after extension/PCR reactions. One of the 13 A-T-Ps (obtained from Strain 1) only produced an amplification product with the genomic DNA from which it was generated. This indicated the captured sequence is unique to Strain 1; that is, the captured sequence is not present in other strains, or that the anchor primer sequence in other strains is not close enough to the captured sequence for PCR amplification to occur.

In summary, at least nine out of the thirteen A-T-P prepared were informative and functional in extension/PCR. This represents a 70% success rate in the preparation of informative and functional A-T-P

5. Extension/PCR of A-T-Ps on genomic DNA of different bacterial strains

Extension/PCR reactions were carried out with uncut S. aureus genomic DNA from the 4 strains using the 9 functional A-T-Ps. Extension reactions with different A-T-Ps were set up as described in the Table 26 and heated at 94°C for 2 minutes. The extension reactions were repeated 5 times under the following temperature profiles in a thermocycler: Thirty seconds at 94°C, 30 seconds at 37°C and 5 minutes at 72°C.

Table 26. Extension of A-T-P on *S. aureus* genomic DNA

	Sample RXN	
Components	Volume (µl)	Final Conc.
A-T-P (30-50 fmol/µl)	1.7	2.5-4.25 fmol/µl
S. aureus genomic DNA (50 ng/µl)	1	2.5 ng/µl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	2	1X
$MgCl_2$ (25 mM)	1.2	1.5 mM
dNTPs (2mM each)	2	200 µM
Taq DNA Polymerase (5 U/µl)	1	0.1 U/µl
ddH_2O	11.1	
	20	

After the extension step, the PSS2 primer and the 3'IS431 or the 3'DEGIS431 primer were added to each extension reaction. The second step of the extension/PCR reaction (Table 27) was carried out under the following conditions: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3 minutes; 72 minutes for 5 minutes.

Table 27. PCR amplification of the extension reaction

Components	Volume (µl)	Final Concentration
Extension reaction	20	
PSS2 (25 pmol/µl)	1	0.5 pmol/µl
3'IS431 primer or 3'DEGIS431 primer (25 pmol/µl)	1	0.5 pmol/µl
10X PCR Buffer (100mM Tris-HCl (pH8.8), 500mM KCl and 0.8% Nonidet P40)	3	1X
$MgCl_2$ (25 mM)	1.8	1.5 mM
dNTPs (2mM each)	3	200 µM
ddH_2O	20.2	
	50	

Results are shown in Table 28. A-T-P13 and A-T-P14 were redundant as they were prepared from the same restriction cut library (HindIII) but from different strains (2 and 1, respectively). The lengths of the amplified products from extension/PCR reactions were the same further indicated that they were redundant; that is, they have the same captured sequence. This condition also applied for A-T-P10 and A-T-P11, and A-T-P8 and A-T-P3. The patterns of amplification products obtained from extension/PCR for the four strains can be classified into 4 groups.

Classification of patterns of amplified products obtained from extension/PCR reactions

Group 1 pattern, obtained from A-T-P2, A-T-P7, A-T-P13, and A-T-P14, showed that Strain 1 and Strain 4 have common amplified products, and Strain 2 and Strain 3 have another common amplified products. As the PCR products obtained from the *in vitro* PstI libraries were sequenced, they were found to correspond to the mec DNA region. Therefore, the A-T-Ps prepared from them should amplify fragments from the mec DNA region. From the GenBank sequence of mec DNA, the PagI and HindIII sites can be determined so that the correct sizes of the extension/PCR products from the corresponding A-T-Ps can be predicted. Indeed, there are a PagI site about 100 bp in front of the PstI site, and a HindIII site about 200 bp behind the PstI site. Thus this information matched the experimental results where the amplification product of

Table 28. Extension/PCR Products from Uncut *S. aureus* genomic DNA

No	A-T-P	3'IS431 (SS)				3'DEGIS431 (DE)			
		Strain 1	Strain 2	Strain 3	Strain 4	Strain 1	Strain 2	Strain 3	Strain 4
2	S1PSTI20	2000 bp	900 bp	900 bp	2000 bp	1750 bp	650 bp	650 bp	1750 bp
7	S2PAGI8	1900 bp	800 bp	800 bp	1900 bp	1650 bp	550 bp	550 bp	1650 bp
13	D2HINDIII8	2200 bp	1100 bp	1100 bp	2200 bp	1900 bp	800 bp	800 bp	1900 bp
14	D1HINDIII19	2200 bp	1100 bp	1100 bp	2200 bp	1900 bp	800 bp	800 bp	1900 bp
6	S2PAGI20	-	2000 bp	2000 bp	-	-	1750 bp	1750 bp	-
10	D4ECORI11	-	-	1400 bp	1400 bp	-	-	1150 bp	1150 bp
11	D3ECORI11	-	-	1400 bp	1400 bp	-	-	1150 bp	1150 bp
8	S3PAGI8	800 bp	-	800 bp	-	550 bp	-	550 bp	-
3	S1PAGI8	800 bp	-	800 bp	-	550 bp	-	550 bp	-

A-T-P-7 was 100 bp smaller than that of A-T-P-2, and the amplification products of A-T-P13 and A-T-P14 were 200 bp longer than that of A-T-P2.

Group 2 pattern, obtained from A-T-P 6, showed that only Strain 2 and Strain 3 have amplified products of the same length. The absence of amplified products for Strain 1 and Strain 4 could mean that the captured sequence is not present in these strains, or the IS431 element is not inserted close by the captured sequence and/or in the correct orientation.

Group 3 pattern, obtained from A-T-P10 and A-T-P 11, showed that only Strain 3 and Strain 4 have amplified products of the same length.

Group 4 pattern, obtained from A-T-P8 and A-T-P 3, showed that only Strain 1 and Strain 3 have amplified products of the same length.

In conclusion, we have isolated unique sequences that enabled us to classify (or type) the 4 strains into 4 groups. The PCR products from which these A-T-Ps were prepared were sequenced. The captured sequences of each A-T-P were determined, and the sequence homologies of the PCR products to known sequences were found by searching the GenBank using the BLAST program (NIH web site). The results are summarized in Table 29.

Summary of the bacterial gene region mapped with the A-T-Ps

Group 1 pattern represents the mec DNA region as discussed earlier. Group 2 pattern turned out to represent a second known IS431 insertion site further downstream from the first IS431 site on the mec DNA region. It is interesting to note that only Strain 2 and Strain 3 have this second insertion site. Group 3 pattern was mapped to a *Staphylococcus* multi-resistance plasmid DNA (pSK1) known to be associated with trimethoprium resistance and has 3 copies of IS257 elements (GenBank Accession #X13290). Only Strain 3 and Strain 4 have this plasmid sequence. Group 4 pattern was mapped to a Staphylococcus conjugative plasmid DNA (pSK41) known to have multiple copies of IS257 elements. Only Strain 1 and Strain 3 have this plasmid sequence.

Figures 7-9 summarize the locations of the PCR products (in Table 25) obtained from the different libraries and the different strains with respect to the four DNA regions mapped. The locations were deduced from A-T-P extension/PCR experiments, and sequencing data obtained from PCR products produced from the libraries. PCR products obtained with the specific 3'IS431 anchor primer and the degenerate 3'DegIS431 anchor primer are indicated as SS and DE, respectively. The PCR products from which A-T-Ps were prepared are indicated with broad arrows in these Figures.

Figure 7 shows the mec DNA region with the 2 copies of IS431 elements as documented in the GenBank sequence. In Figure 7(a), the PCR products obtained with PagI and EcoRI libraries are shown. All 4 strains have the first IS431 element albeit the IS431 elements of Strain 2 and Strain 3 are inserted closer to the PagI site. There is no EcoRI site close by the first IS431 element so PCR products for EcoRI *in vitro* libraries were not obtained for this region.

Table 29. A-T-P generated from PCR products obtained from *in vitro* libraries of different strains of *S. aureus*

No	Name	A-T-P Captured Sequence	Strain	Library	Sequence information obtained	Sequence mapped to
2	S1PSTI20	CTGCAGAAGCATCTTT (SEQ ID NO: 16)	1	Pst I	~1500 bp	Mec DNA Region 1
7	S2PAGI8	TCATGAATGATAATCA (SEQ ID NO: 17)	2	Pag I	~750 bp	
13	D2HINDIII8	AAGCTTGAAATGAGCA (SEQ ID NO: 18)	2	Hind III	~800 bp	
14	D1HINDIII19	N.D.	1	Hind III	-	
6	S2PAGI20	TCATGAGTACTGATAT (SEQ ID NO: 19)	2	Pag I	~750 bp	Mec DNA Region 2
10	D4ECORI11	GAATTCACCCACCTCA (SEQ ID NO: 20)	4	Eco RI	~850 bp	pSK 1 DNA
11	D3ECORI11	N.D.	3	Eco RI	-	
8	S3PAGI8	TCATGATAAAGAATCT (SEQ ID NO: 21)	3	Pag I	~750 bp	pSK41 DNA
3	S1PAGI8	N.D.	1	Pag I	-	

N.D. – Not Determined

SEQ ID NOS: 22 to 27 represent the sequence of the A-T-P including the captured sequence of SEQ ID NOS: 16 to 21, respectively.

In Figure 7(b), the PCR products obtained with HindIII and PstI libraries are shown. With HindIII libraries using the SS anchor primer in PCR reactions, all of the strains produced a 200 bp product because there is a HindIII site (at position 7013) just in front of the anchor primer. These PCR products are not useful for diagnostic purposes and are not shown in this Figure. But they are useful as positive controls in our experiments. With HindIII libraries using the DE anchor primer in PCR reactions, PCR products were obtained for all 4 strains as expected from the DNA sequence. Again the PCR products from Strain 2 and Strain 3 are shorter than the PCR products of Strain 1 and Strain 4. With PstI libraries, the expected PCR products from each strain were obtained using both the SS and DE anchor primers in PCR reactions. These results confirmed that all of the 4 strains have the first IS431 element in the mec DNA region.

For the second IS431 element in the mec DNA region, the A-T-P extension/PCR results suggested that only Strain 2 and Strain 3 have this region. With PCR reactions of EcoRI libraries, both Strain 2 and Strain 3 produced similar size products (Figure 7(a)). With PCR reactions of PagI libraries, only Strain 2 produced an amplified product and not Strain 3. However, an extension/PCR product was obtained for Strain 3 when the A-T-P (S2PagI20) prepared from the Strain 2 PCR product was used (See Table 28).

In Figure 8, the pSK1 plasmid DNA containing 3 copies of IS257 elements is shown. There are no PstI or PagI sites in the sequence. With EcoRI libraries in PCR reactions, only Strain 3 and Strain 4 produced amplified products of similar lengths (1400 bp). From extension/PCR reactions using A-T-P (D4ECORI10) prepared from Strain 4, only genomic DNA from Strain 3 and Strain 4 produced amplified products of 1400 bp long (Table 28). Thus this confirmed that only Strain 3 and Strain 4 have the captured sequence, which is in the pSK1 plasmid DNA.

In Figure 9, the pSK41 plasmid DNA containing 7 copies of IS257 elements is shown. There are no PstI sites in the pSK41 DNA. Figure 9(a) shows the left-halves of the pSK41 in close-up. A-T-P (S3PagI8) prepared from the PCR product from the PagI library of Strain 3 was instrumental in the identification of this plasmid DNA in extension/PCR

reactions. Only Strain 1 and Strain 3 produced 800 bp products in extension/PCR reactions (Table 28). The DNA sequence of the 800 bp PCR product matched the region in front of the first IS257 element in pSK41 DNA. From the known pSK41 DNA sequence documented in GenBank, PCR products of ~860 bp should also be obtained from PagI libraries of Strain 1 and Strain 3. These bands would be so close to the 800 bp products that we would not be able to separate them on agarose gels used in the experiments. In fact, the 800 bp PCR products from Strain 1 and Strain 3 (PagI libraries, Table 25) were determined to be comprised of multiple sequences by running the sample in higher percentage agarose gels (before A-T-P was prepared). From sequencing the "non-purified" 800 bp product of Strain 1, it was found that the sequencing data was not readable suggesting that the product comprised of multiple sequences.

With EcoRI libraries in PCR reactions, only Strain 1 produced a 4000 bp product. A-T-P (S1EcoRI40) was prepared from this PCR product. When this A-T-P was used in extension/PCR reactions, only the genomic DNA of Strain 1 produced a similar sized band. This DNA product was not sequenced, but the size (4000 bp) matched what would be expected if it comes from the region in front of the third IS257 element of pSK41 DNA.

Figure 9(b) shows the right-halves of pSK41 DNA. The 2500 bp PCR product obtained from the HindIII library of Strain 1 was tentatively assigned to this region. The product size matches the region in front of the fifth IS257 element of pSK41 DNA. A-T-P (D1HindIII25) was prepared from this PCR product, but the extension/PCR reactions of the 4 strains produced only faint bands and made it hard to interpret the results.

Table 30 summarizes the information obtained from these experiments on the four strains of *S. aureus*. Basically, these 4 S. aureus strains can be distinguished using only 4 A-T-Ps (that is, four captured sequences). As the sequences of the four variable regions have been determined, optimized PCR primers for multiplex PCR can be prepared and the strains can be distinguished using multiplex PCR with the four optimized PCR primers.

Example 4. Use of A-T-Ps in bacterial strain typing

Twelve different *S. aureus* strains isolated from different outbreaks at different hospitals were obtained from NIH. To do this screening without bias, the details of the strains were not supplied to us. We did know that one of the strains is methicillin-sensitive. As a preliminary screen, four of the A-T-Ps (#2, 6, 10, and 8 that distinguish the four different genetic locations) were used in extension/PCR reactions with the genomic DNAs of the 12 new *S. aureus* strains. The extension/PCR reactions were carried out in exactly the same way as described in step 5 of Example 3.

The results obtained with the 3'IS431 (SS) anchor primer and 3'DEGIS431 (DE) anchor primer are shown in Table 31 and Table 32, respectively. From the A-T-P extension/PCR results with the SS anchor primer, the 16 strains of S. aureus can be assigned into 11 groups. With the exception of Group E (6 members), every group only has 1 member. For Group E, the mec DNA region1 was the only DNA region detected by extension/PCR reactions. Therefore, new additional A-T-P would be needed to further sub-classify these strains; additional A-T-P could be prepared from *in vitro* libraries of this group.

Nevertheless, the resolving power of using just 4 A-T-Ps was demonstrated. If we only consider the presence or absence of a band, the combination of 4 A-T-Ps should give us a resolving power of 16 groups (2^4). With the possibility of the product band being different sizes, the resolving power will be greater than 16 groups. We were able to distinguish 11 groups; variations in the band sizes were observed for the amplified products from mec DNA region 1 and pSK41 DNA. Strain 16 is a methicillin-sensitive *S. aureus* strain, and should not have any insertion elements associated with methicillin-resistance. Indeed, we did not obtain any amplified products for this strain.

When the DE anchor primer was used instead of the SS anchor primer, amplified products were obtained for additional strains in the mec DNA region 1 and region 2. Amplified products with the similar size as those of Strain 1 and Strain 4 were obtained for

Strain 12 and Strain 15 when DE anchor primer was used in the extension/PCR reactions with A-T-P(S1PSTI20). Likewise, additional amplified products were obtained for Strains 5, 7, 8, 10, 12, 14 and 15 when DE

Table 30. Summary of results for the *S. aureus* strains with respect to the different DNA regions mapped

Strain	Mec DNA		pSK1 DNA	pSK41 DNA
	Region 1	Region2		
1	Present	Not Present	Not Present	Present
2	Present (with an internal 1100 bp deletion)	Present	Not Present	Not Present
3	Present (with an internal 1100 bp deletion)	Present	Present	Present
4	Present	Not Present	Present	Not Present

Table 31. Screening results for the 12 new strains of *S. aureus* with the A-T-Ps for four different genetic locations and SS anchor primer

Strain	S1PSTI20 (Mec DNA Region 1)	S2PAGI20 (Mec DNA Region 2)	D4ECORI11 (pSK1 DNA)	S3PAGI8 (pSK41 DNA)	Group Assignment
1	+	-	-	+	A
2	+ (with deletion)	+	-	-	B
3	+ (with deletion)	+	+	+	C
4	+	-	+	-	D
5	+ (similar size as strains 2 & 3)	-	-	-	E
6	+ (similar size as strains 2 & 3)	-	+	+ (Larger Band)	F
7	+ (similar size as strains 2 & 3)	-	-	-	E
8	+ (similar size as strains 2 & 3)	-	-	-	E
9	+ (similar size as strains 2 & 3)	-	-	-	E
10	+ (similar size as strains 2 & 3)	-	-	-	E
11	+ (similar size as strains 2 & 3)	+	-	+ (Smaller Band)	G
12	-	-	+	+	H
13	+ (similar size as strains 2 & 3)	-	+	+	I
14	+ (similar size as strains 2 & 3)	-	-	-	E
15	-	-	-	+ (Larger Band)	J
16	-	-	-	-	K

Table 32. Screening results for the 12 new strains of *S. aureus* with the A-T-Ps for four different genetic locations and DE anchor primer

Strain	S1PSTI20 (Mec DNA Region 1)	S2PAGI20 (Mec DNA Region 2)	D4ECORI11 (pSK1 DNA)	S3PAGI8 (pSK41 DNA)	Group Assignment
1	+	-	-	+	A
2	+ (with deletion)	+	-	-	B
3	+ (with deletion)	+	+	+	C
4	+	-	+	-	D
5	+ (similar size as strains 2 & 3)	+	-	-	B
6	+ (similar size as strains 2 & 3)	-	+	+ (Larger Band)	F
7	+ (similar size as strains 2 & 3)	+	-	-	B
8	+ (similar size as strains 2 & 3)	+	-	-	B
9	+ (similar size as strains 2 & 3)	-	-	-	E
10	+ (similar size as strains 2 & 3)	+	-	-	B
11	+ (similar size as strains 2 & 3)	+	-	+ (Smaller Band)	G
12	+	+	+	+	H
13	+ (similar size as strains 2 & 3)	-	+	+	I
14	+ (similar size as strains 2 & 3)	+	-	-	B
15	+	+	-	+ (Larger Band)	J
16	-	-	-	-	K

anchor primer was used in the extension/PCR reactions with A-T-P(S2PAGI20). At this time, it is not known what is the exact reason that additional bands were obtained for these strains. One possibility is that the transposase gene sequences of the insertion elements in these strains are slightly different than the sequence of the IS431 anchor primer. Therefore, only the degenerate anchor primer (3'DEGIS431) can hybridize to the DNA of these strains to give amplified products.

Even with the additional amplified products, the screening result was not that different from the result obtained with the SS anchor primer. The strains screened can still be classified into 11 groups. Five members (Strains 5, 7, 8, 10, and 14) of the previously classified Group E are now re-classified to Group B. Strain 9 becomes the lone member in Group E. None of the members of the other groups changed.

In summary, it was demonstrated that different S. aureus strains can be distinguished (typed) with a set of 4 A-T-Ps prepared from 4 independent strains. These results showed that unique sequences that can be used to amplify polymorphic regions in bacteria were easily isolated by the SCOT method. By using the present invention, polymorphic bacterial regions can easily be identified. Optimized PCR primers may be developed from the sequence information obtained from the amplified products from the extension/PCR reactions of A-T-Ps. Database may be created to document the amplification products from different bacterial strains using the optimized PCR primers. Such "fingerprint" database could be used to identify bacterial strains in future outbreaks of bacterial infections.

I CLAIM:

1. A single stranded oligonucleotide capable of forming a partially double stranded structure having a closed end and an open end, the structure comprising at the closed end a first double stranded region, at the open end, a second double stranded region, between the first and second double stranded regions, a region of non-complementary sequences wherein the non-complementary sequences comprise a known sequence to which a primer can specifically hybridize, and the second double stranded region comprises a class IIS RE recognition site.

2. The oligonucleotide according to claim 1 wherein the first and second double stranded regions are GC rich regions.

3. The oligonucleotide according to claim 2 having the sequence shown in SEQ ID No.1.

4. The oligonucleotide according to claim 3 having the sequence selected from the group consisting of SEQ ID No.2, SEQ ID NO; 3, SEQ ID NO: 4 and SEQ ID NO: 5.

5. A primer selected from the group consisting of SEQ ID No.16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20 and SEQ ID NO: 21, any complement, optimized or modified derivative thereof.

6. A primer having the sequence selected from the group consisting of SEQ ID NO.22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 and SEQ ID NO: 27, any optimized or modified derivative thereof.

7. A method of analyzing a DNA sample comprising the steps of:

(a) ligating a first oligonucleotide and a DNA fragment to form a ligated product wherein the oligonucleotide comprises a known sequence and a recognition site of a

class IIS RE;

(b) digesting the ligated product with the class IIS RE to release a second oligonucleotide comprising the first oligonucleotide and a sequence from the DNA fragment;

(c) forming a reaction mixture comprising the DNA sample, the second oligonucleotide, a first amplification primer and a second amplification primer under conditions suitable to drive extension and amplification wherein the first amplification primer specifically hybridizes to the known sequence and the second amplification primer specifically hybridizes to a sequence in the DNA fragment;

(d) determining the presence of an extension product in the reaction mixture.

8. The method according to claim 7 wherein the first oligonucleotide is an oligonucleotide according to claim 1.

9. The method according to claim 8 further comprising the step of sequencing the extension product.

10. The method according to claim 8 further comprising the step of amplifying the ligated product with the first and second amplification primers prior to step (b).

11. The method according to claim 8 further comprising the step of purifying the ligated product after step (a).

12. The method according to claim 11 wherein the ligated product is purified with a DNA extraction kit.

13. A method of preparing an extension primer comprising the steps of:

(a) digesting genomic DNA of a species to generate DNA fragments;

(b) ligating the resulting DNA fragments with a first oligonucleotide to form ligated products wherein the oligonucleotide comprises a known sequence and a recognition site for a class IIS RE;

(c) amplifying one or more ligated products with a first and second primer wherein the first primer specifically hybridizes to the known sequence and the second primer specifically hybridizes to a sequence in the genomic DNA;

(d) digesting the one or more amplified product with the class IIS RE to generate an extension primer comprising the first oligonucleotide and a sequence from the genomic DNA.

14. The method according to claim 13 wherein the first oligonucleotide I is an oligonucleotide according to claim 1.

15. The method according to claim 14 further comprising the step of purifying the genomic DNA after step (a).

16. The method according to claim 15 further comprising comparing the one or more amplified product to one or more amplified product obtained with genomic DNA of another member of the species prior to step (d).

17. The method according to claim 16 wherein the one or more amplified products are compared by agarose gel electrophoresis.

18. The method according to claim 16 further comprising the step of purifying ~e one or more amplified product prior to step (d).

19. The method according to claim 18 wherein the one or more amplified product is purified by agarose gel electrophoresis.

20. The method according to claim 14 further comprising the step of purifying the

extension primer.

21. The method according to claim 20 wherein the extension primer is purified by agarose gel electrophoresis.

22. The method according to claim 14 further comprising the step of amplifying the extension primer.

23. The method according to claim 22 wherein the step of amplifying the extension primer comprises the steps of ligating the extension primer to a second oligonucleotide comprising a known sequence to form a second ligated product; amplifying the second ligated product with a first and second primer which specifically hybridize to the known sequence in the first and second oligonucleotide, respectively and digesting the amplified second ligated product with the class IIS RE.

24. The method according to claim 22 further comprising the step of reversing functional directionality of the extension primer.

25. The method according to claim 24 wherein the step of reversing functional directionality comprises the steps of ligating the extension primer to a second oligonucleotide comprising a known sequence and a recognition site for a class IIS RE that differs from the class IIS RE recognition site on the first oligonucleotide and digesting with a class IIS RE that recognizes the recognition site on the second oligonucleotide .

26. The method according to claim 25 wherein the class IIS RE that recognizes the recognition site on the second oligonucleotide cleaves the second oligonucleotide at the same distance outside the recognition site as the class IIS RE that recognizes the class IIS RE recognition site on the first oligonucleotide.

27. The method according to claim 26 wherein the class IIS RE that recognizes the recognition site on the first oligonucleotide is GsuI and the class IIS RE that recognizes the recognition site on the second oligonucleotide is Eco57I.

28. The method of claim 14 wherein the species is a bacterial species.

29. The method of claim 28 wherein the second amplification primer specifically hybridizes to an insertion element.

30. The method of claim 29 wherein the genomic DNA of the bacterial species is digested with a restriction enzyme that cuts bacterial genome at a frequency of once per about 2000 to 8000 base pairs.

31. The method of claim 29 wherein the bacterial DNA is digested with PstI.

32. The method of claim 31 wherein the first oligonucleotide has the sequence shown in SEQ ID No.2.

33. The method of claim 29 wherein the bacterial DNA is digested with EcoRI.

34. The method of claim 33 wherein the first oligonucleotide has the sequence shown in SEQ ID No.3.

35. The method of claim 29 wherein the bacterial DNA is digested with HindIII.

36. The method of claim 35 wherein the first oligonucleotide has the sequence shown in SEQ ID No.4.

37. The method of claim 29 wherein the bacterial DNA is digested with PagI.

38. The method of claim 37 wherein the first oligonucleotide has the sequence shown in SEQ ID No.5.

39. The method of claim 29 further comprising the steps of forming a reaction mixture comprising the genomic DNA of another number of the bacterial species, the extension primer, the first and second primer under conditions suitable to drive extension and

amplification, determining the presence of one or more extension product in the reaction mixture and regenerating the extension primer by digesting the one or more extension product with the class IIS RE used in step (d).

40. A method of typing genomic DNA of a species comprising the steps of:

(a) forming a reaction mixture comprising the genomic DNA, an extension primer, a first amplification primer and a second amplification primer under conditions suitable to drive extension and amplification wherein the extension primer comprises a known sequence and a first sequence of genomic DNA from another member of the species, the first amplification primer specifically hybridizes to the known sequence and the second amplification primer specifically hybridizes to a second sequence downstream from the first sequence of the genomic DNA of said another member of the species;

(b) determining the presence of an extension product in the reaction mixture.

41. The method according to claim 40 wherein the species is a bacterial species .

42. The method according to claim 41 wherein the second sequence is an insertion element.

43. A kit for preparing an extension primer comprising:

(a) an oligonucleotide according to claim 1;

(b) a primer which can specifically hybridize to the known sequence of the oligonucleotide .

44. The kit according to claim 43 further comprising instruction for preparing an extension primer.

45. The kit according to claim 44 further comprising an adaptor for amplification of

the extension primer wherein the adaptor comprises a known sequence to which a primer can specifically hybridize and class IIS RE recognition site that is different than the class IIS recognition site on the oligonucleotide.

46. The kit according to claim 45 further comprising a primer which can specifically hybridize to the known sequence on the adaptor.

47. The kit according to claim 46 further comprising two class IIS restriction enzymes which recognize the class IIS RE recognition sites on the oligonucleotide and the adaptor, respectively.

48. The kit according to claim 47 wherein the oligonucleotide has a sticky end and the kit further includes a RE that can generate DNA ends complementary to the sticky end.

49. The kit according to claim 48 further comprising a DNA ligase.

50. The kit according to claim 49 wherein the primer which can specifically hybridize to the known sequence on the adaptor is labeled with biotin.

51. The kit according to claim 50 further comprising streptavidin beads.

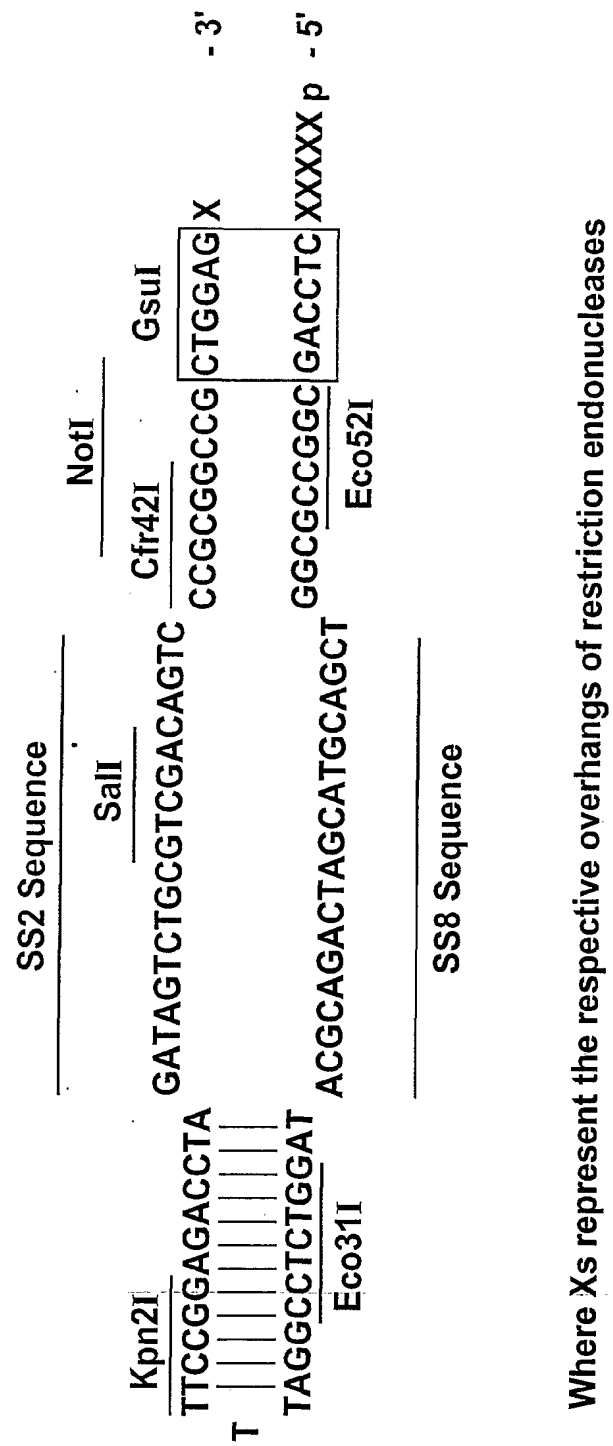

SS2 Sequence
SalI
NotI
Cfr42I
GsuI
Kpn2I
TTCCGGAGACCTA
GATAGTCTGCGTCGACAGTC
CCGCGGCCG
CTGGAG
X
- 3'
T
TAGGCCTCTGGAT
ACGCAGACTAGCATGCAGCT
GGCGCCGGC
GACCTC
XXXX p
- 5'
Eco31I
Eco52I
SS8 Sequence
Where Xs represent the respective overhangs of restriction endonucleases

FIG. 1a

SS4 Sequence
Eco57I
3' -
CAACGTGAGCGATGTCTAGTCTGAAGNN
- 3'
5' -
GTTGCACTCGCTACAGATCAGACTTCp
- 5'

FIG. 1b

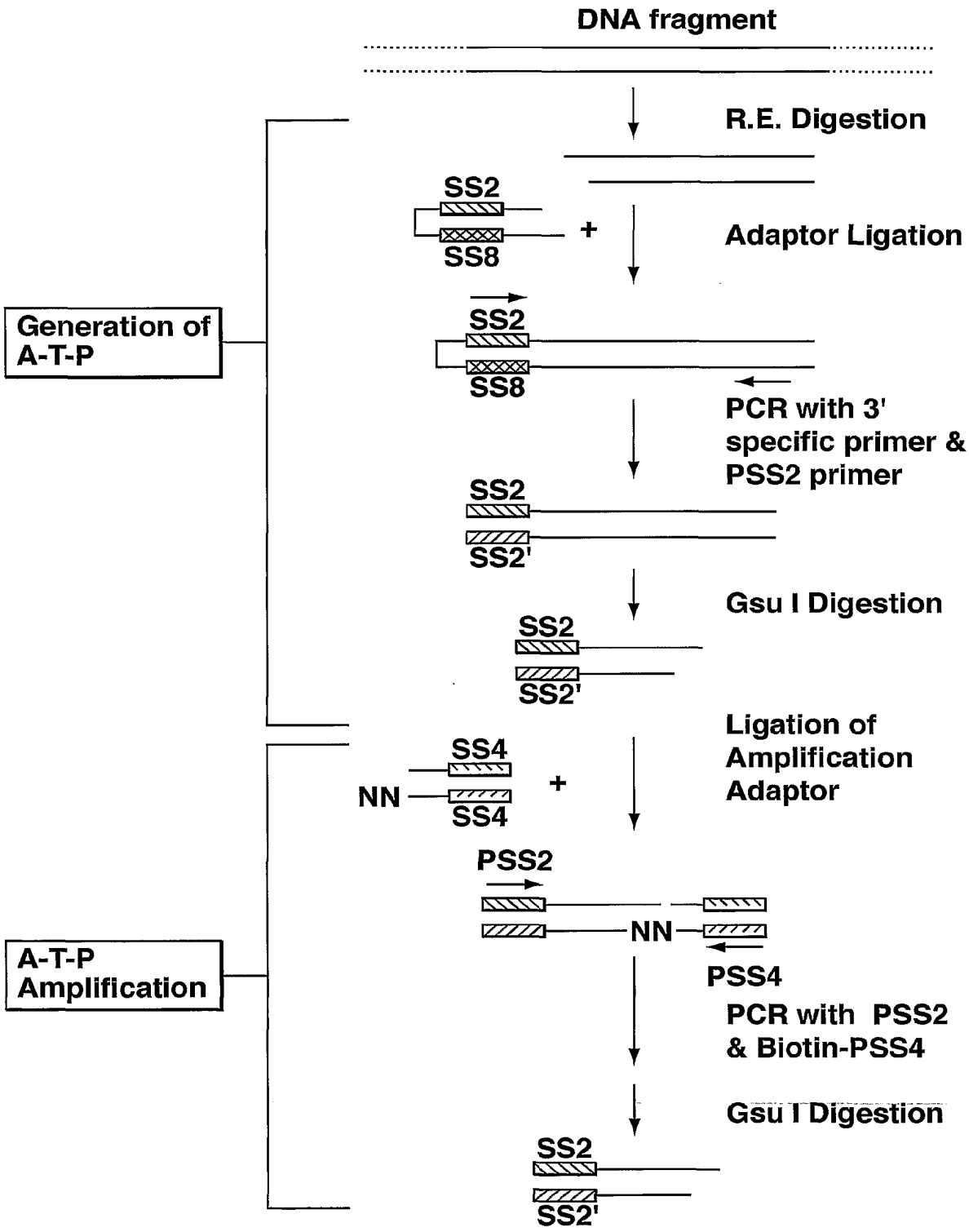

DNA fragment
R.E. Digestion
SS2
SS8
+
Adaptor Ligation
Generation of A-T-P
SS2
SS8
PCR with 3' specific primer & PSS2 primer
SS2
SS2'
Gsu I Digestion
SS2
SS2'
Ligation of Amplification Adaptor
SS4
NN
SS4
+
PSS2
NN
PSS4
A-T-P Amplification
PCR with PSS2 & Biotin-PSS4
Gsu I Digestion
SS2
SS2'

FIG. 2

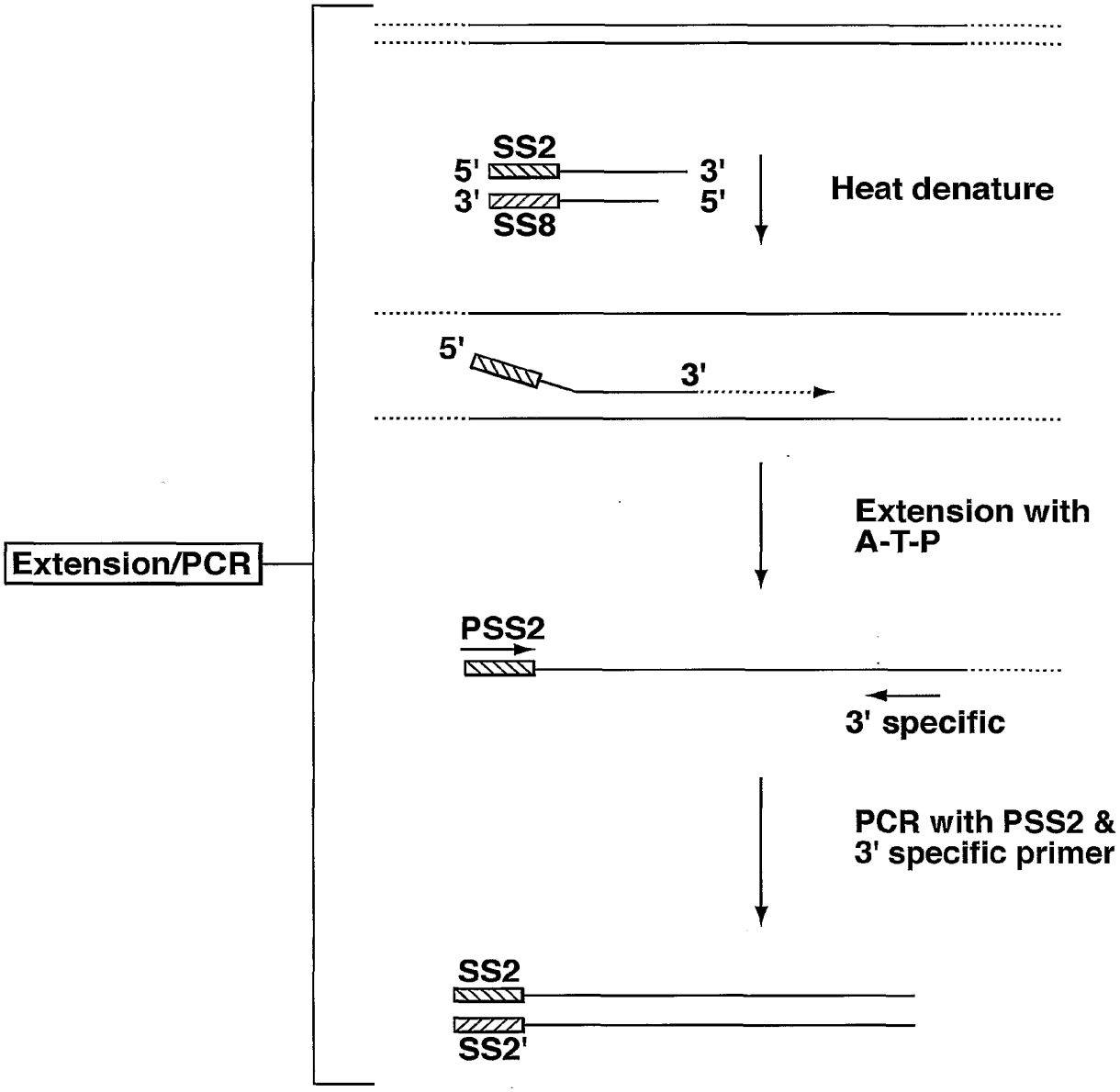

Extension/PCR
SS2
5'
3'
3'
5'
SS8
Heat denature
5'
3'
Extension with
A-T-P
PSS2
3' specific
PCR with PSS2 &
3' specific primer
SS2
SS2'

FIG. 2 (cont'd)

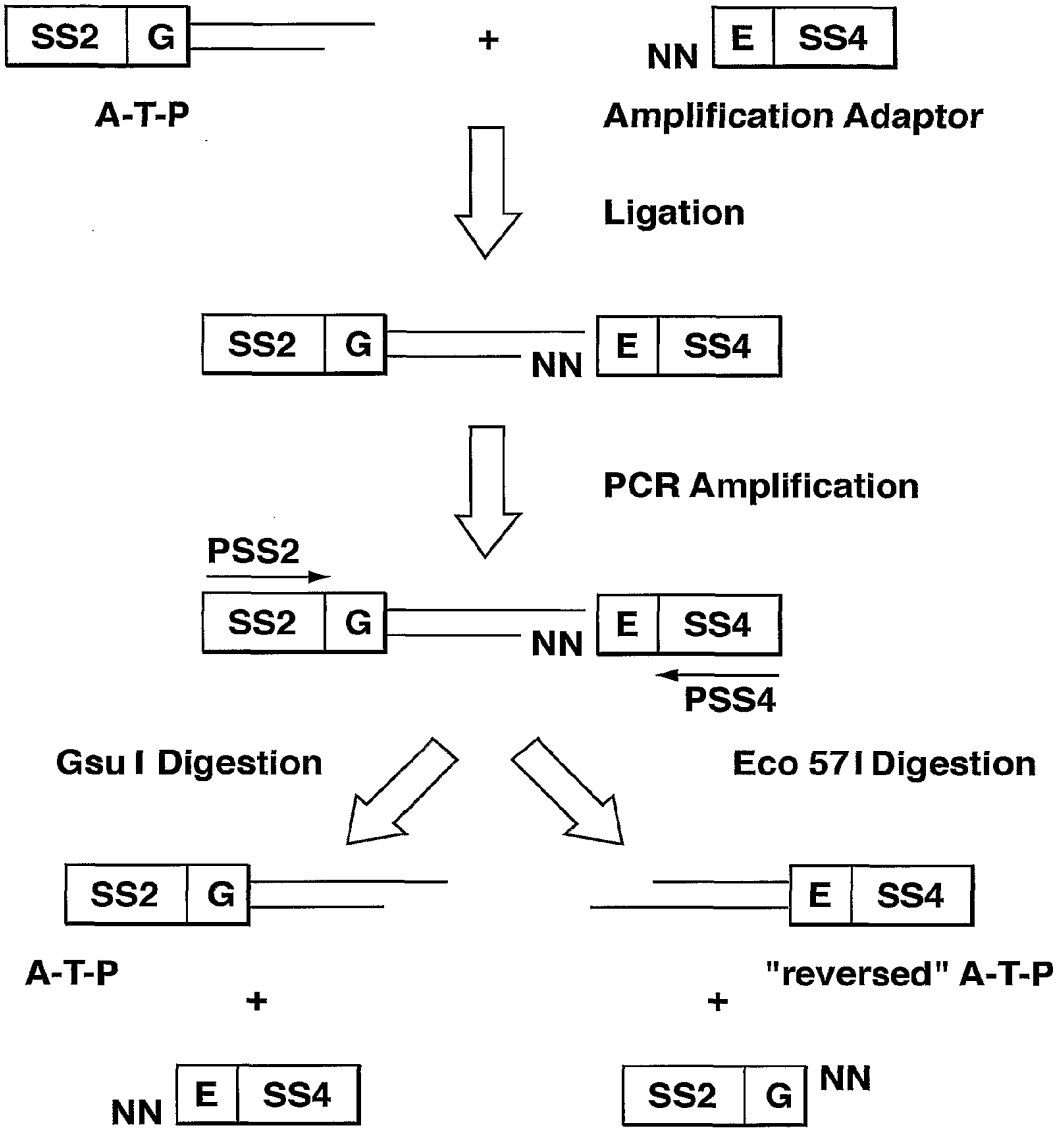

SS2
G
A-T-P
+
NN
E
SS4
Amplification Adaptor
Ligation
PCR Amplification
PSS2
PSS4
Gsu I Digestion
Eco 57I Digestion
"reversed" A-T-P
Legend: SS2,SS4 - Unique sequences
E - Eco 57I recognition site
G - GsuI recognition site
PSS2, PSS4 - PCR primers

FIG. 3

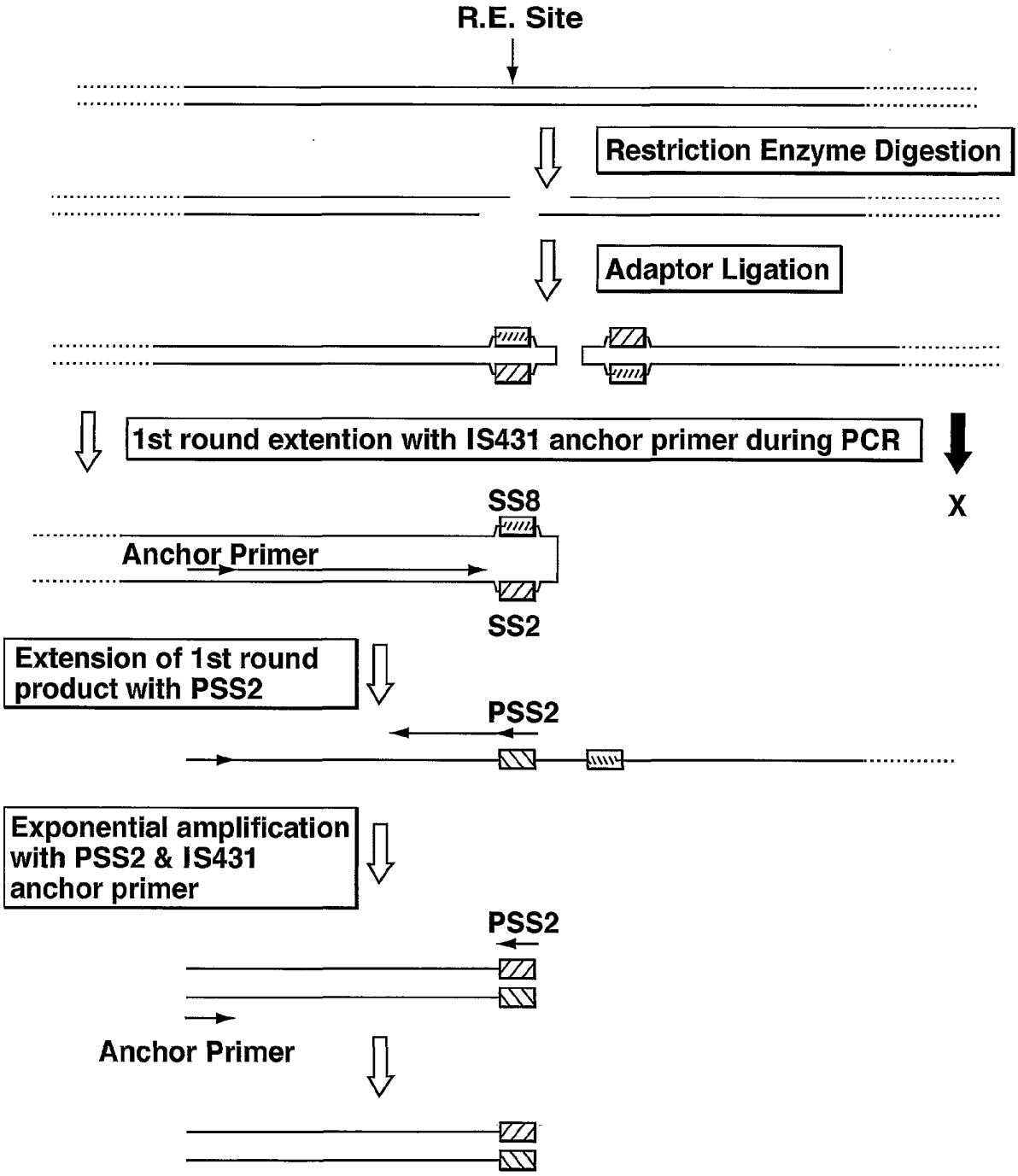

R.E. Site
Restriction Enzyme Digestion
Adaptor Ligation
1st round extention with IS431 anchor primer during PCR
X
SS8
Anchor Primer
SS2
Extension of 1st round product with PSS2
PSS2
Exponential amplification with PSS2 & IS431 anchor primer
PSS2
Anchor Primer

FIG. 4

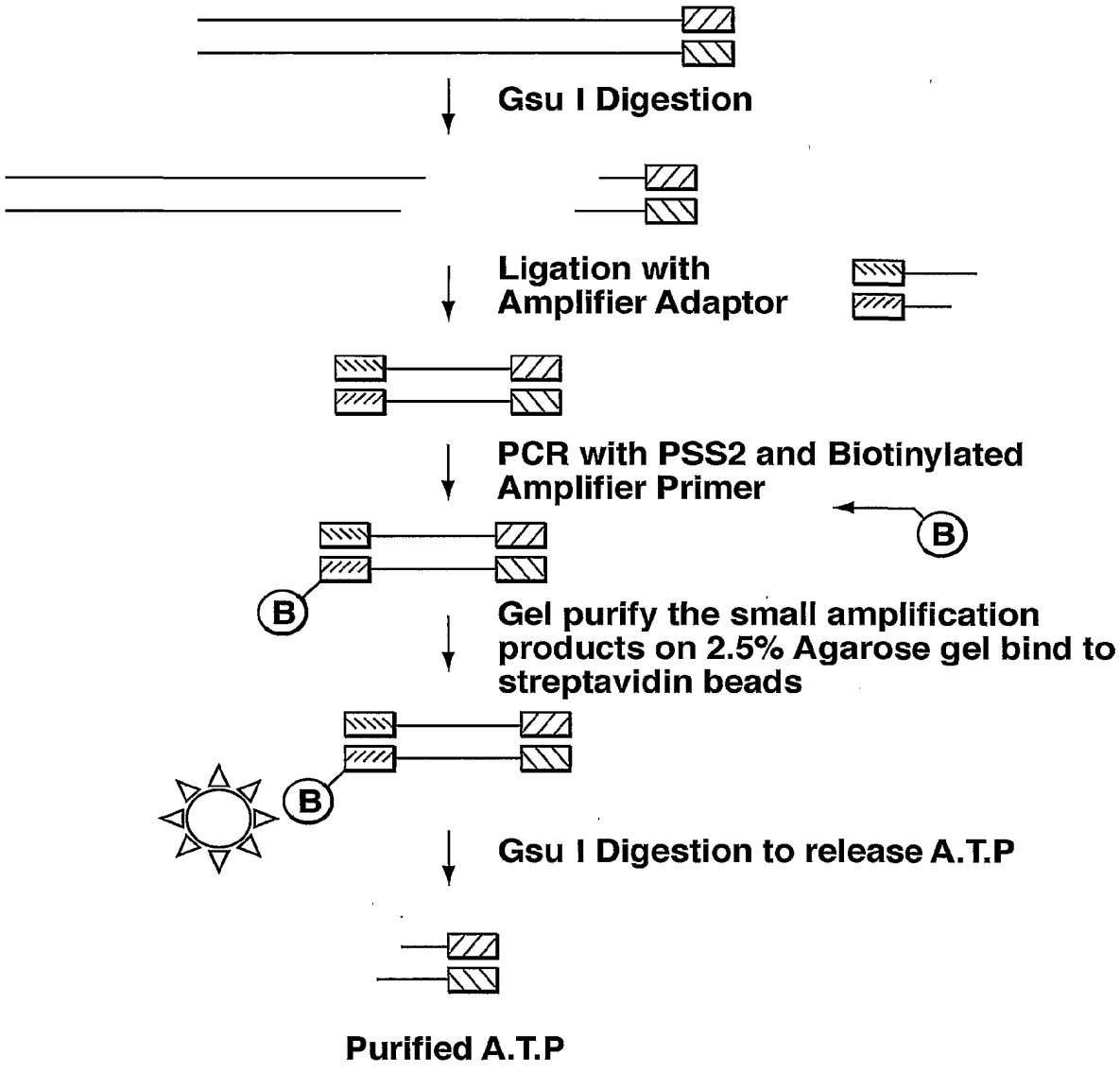

Gsu I Digestion
Ligation with
Amplifier Adaptor
PCR with PSS2 and Biotinylated
Amplifier Primer
B
B
Gel purify the small amplification
products on 2.5% Agarose gel bind to
streptavidin beads
B
Gsu I Digestion to release A.T.P
Purified A.T.P

FIG. 5

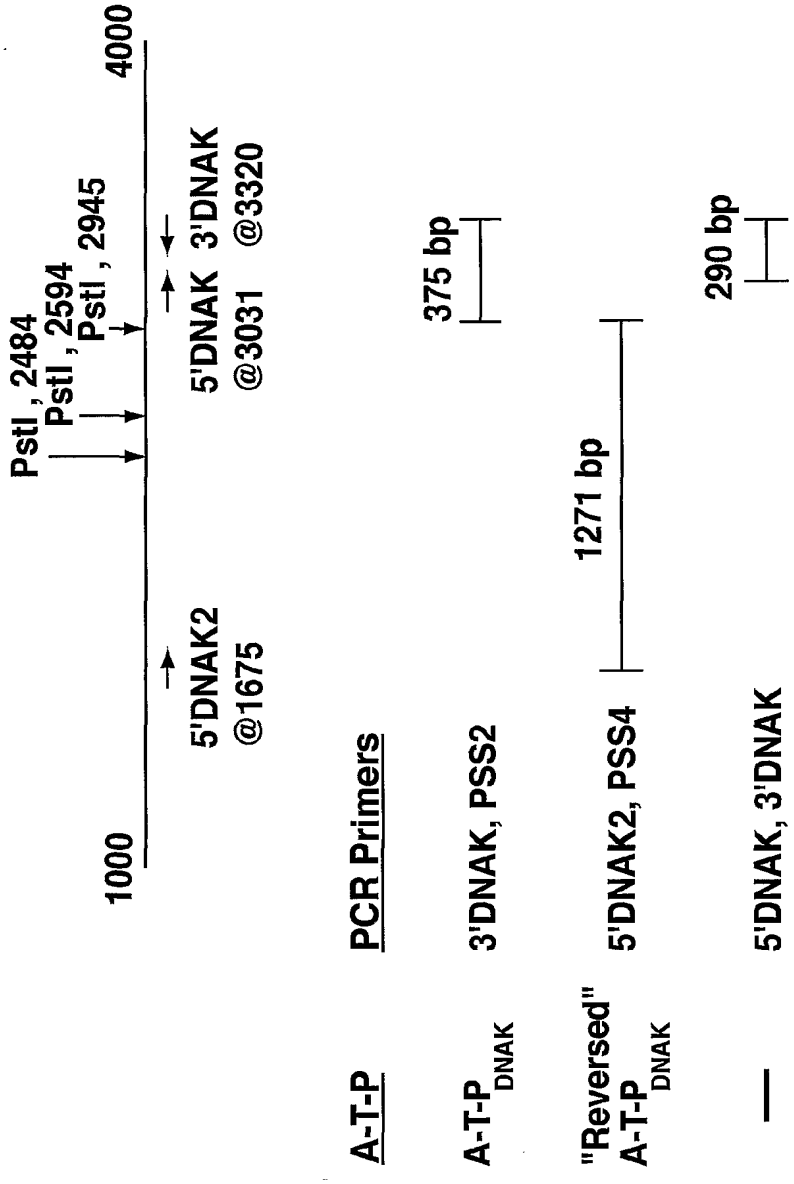

1000
4000
PstI , 2484
PstI , 2594
PstI , 2945
5'DNAK2
@1675
5'DNAK 3'DNAK
@3031 @3320
A-T-P
PCR Primers
A-T-P DNAK
3'DNAK, PSS2
375 bp
"Reversed"
A-T-P DNAK
5'DNAK2, PSS4
1271 bp
5'DNAK, 3'DNAK
290 bp

FIG. 6

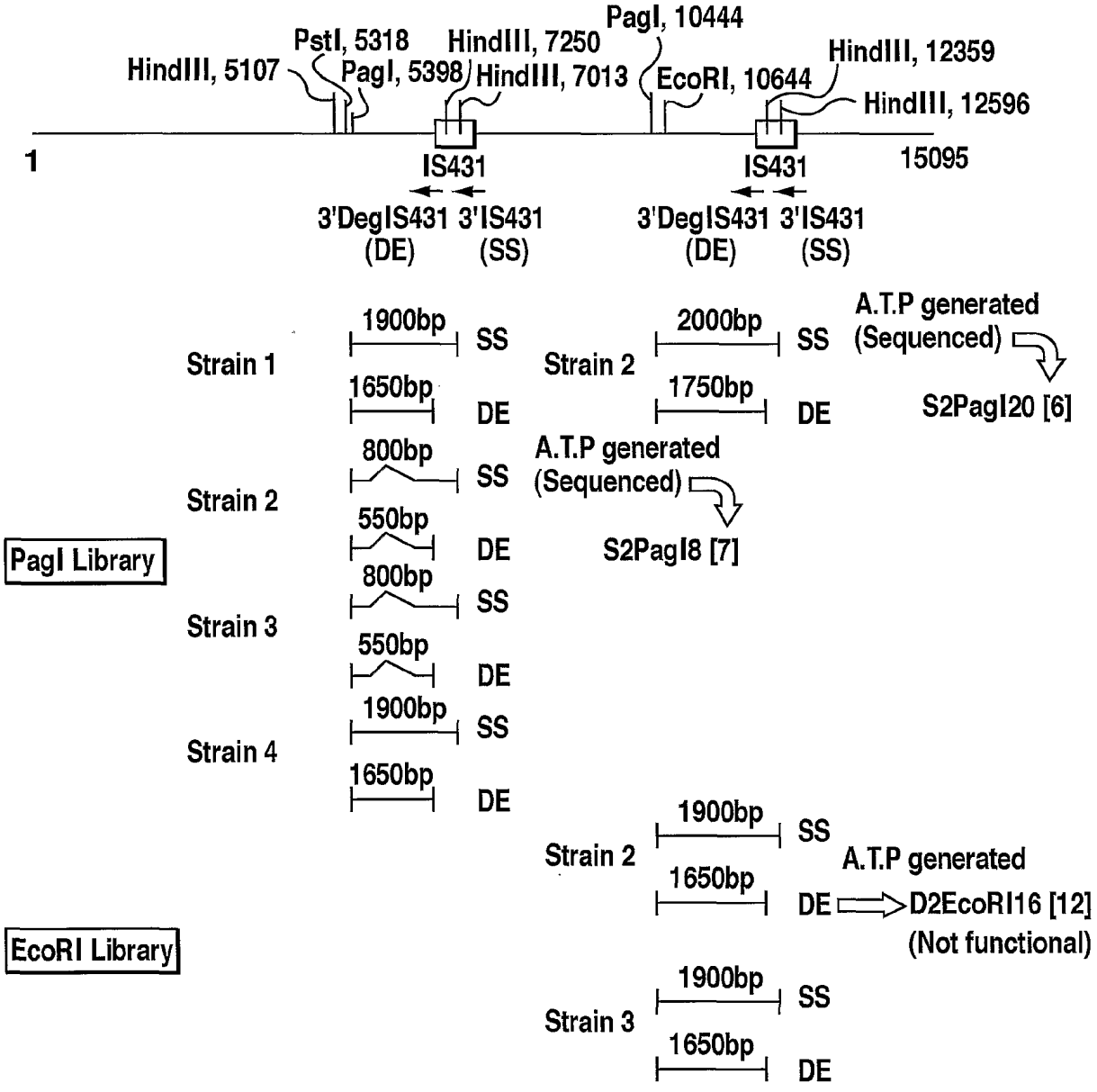

PagI, 10444
PstI, 5318
HindIII, 7250
HindIII, 12359
HindIII, 5107
PagI, 5398
HindIII, 7013
EcoRI, 10644
HindIII, 12596
1
IS431
IS431
15095
3'DegIS431 (DE)
3'IS431 (SS)
3'DegIS431 (DE)
3'IS431 (SS)
Strain 1
1900bp
SS
1650bp
DE
Strain 2
800bp
SS
550bp
DE
Strain 3
800bp
SS
550bp
DE
Strain 4
1900bp
SS
1650bp
DE
PagI Library
Strain 2
2000bp
SS
1750bp
DE
A.T.P generated (Sequenced)
S2PagI20 [6]
A.T.P generated (Sequenced)
S2PagI8 [7]
Strain 2
1900bp
SS
1650bp
DE
A.T.P generated
D2EcoRI16 [12]
(Not functional)
Strain 3
1900bp
SS
1650bp
DE
EcoRI Library

FIG. 7a

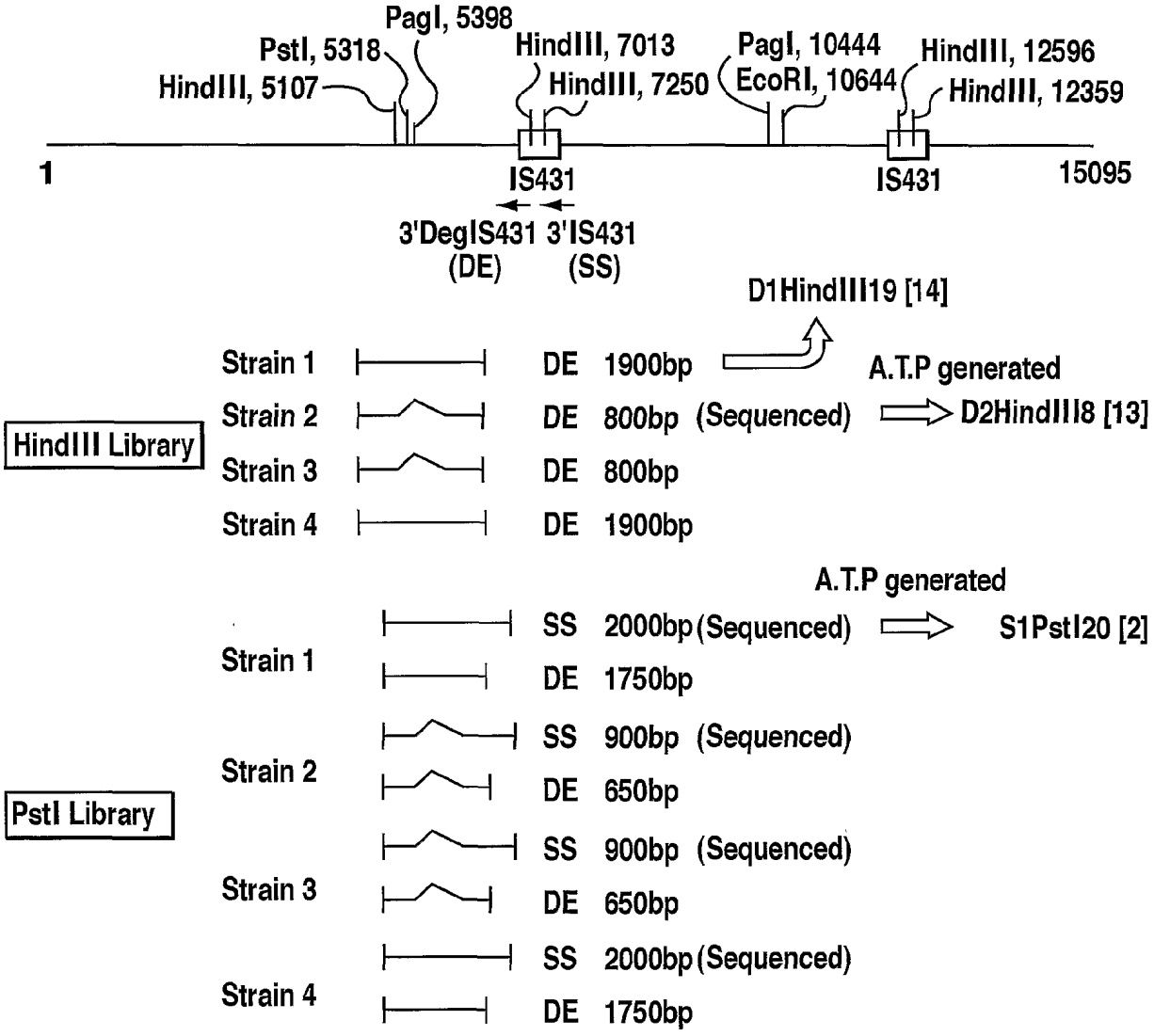

PagI, 5398
PstI, 5318
HindIII, 5107
HindIII, 7013
HindIII, 7250
PagI, 10444
EcoRI, 10644
HindIII, 12596
HindIII, 12359
1
IS431
IS431
15095
3'DegIS431 (DE)
3'IS431 (SS)
D1HindIII19 [14]
HindIII Library
Strain 1
DE 1900bp
A.T.P generated
Strain 2
DE 800bp (Sequenced)
D2HindIII8 [13]
Strain 3
DE 800bp
Strain 4
DE 1900bp
A.T.P generated
PstI Library
Strain 1
SS 2000bp (Sequenced)
S1PstI20 [2]
DE 1750bp
Strain 2
SS 900bp (Sequenced)
DE 650bp
Strain 3
SS 900bp (Sequenced)
DE 650bp
Strain 4
SS 2000bp (Sequenced)
DE 1750bp

FIG. 7b

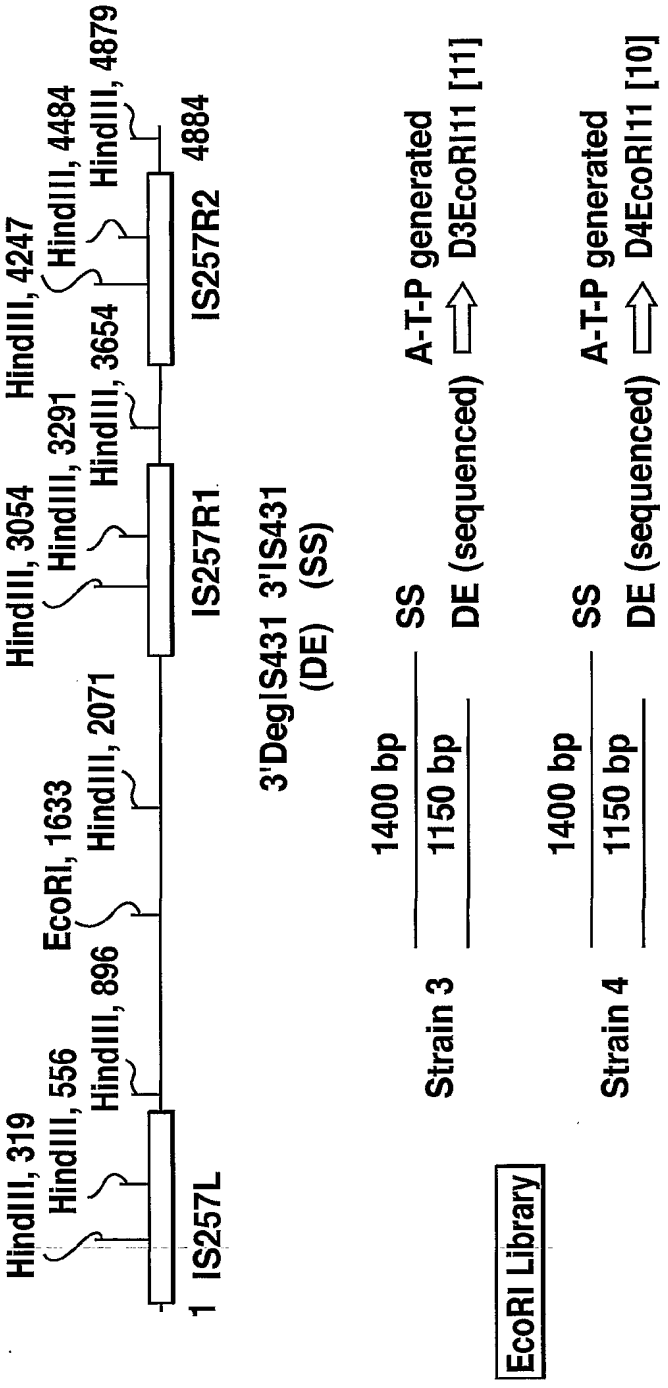

HindIII, 319
HindIII, 556
HindIII, 896
EcoRI, 1633
HindIII, 2071
HindIII, 3054
HindIII, 3291
HindIII, 3654
HindIII, 4247
HindIII, 4484
HindIII, 4879
1 IS257L
3'DegIS431 3'IS431
(DE) (SS)
IS257R1
IS257R2 4884
EcoRI Library
Strain 3
1400 bp
1150 bp
SS
DE (sequenced)
A-T-P generated
D3EcoRI11 [11]
Strain 4
1400 bp
1150 bp
SS
DE (sequenced)
A-T-P generated
D4EcoRI11 [10]

FIG. 8

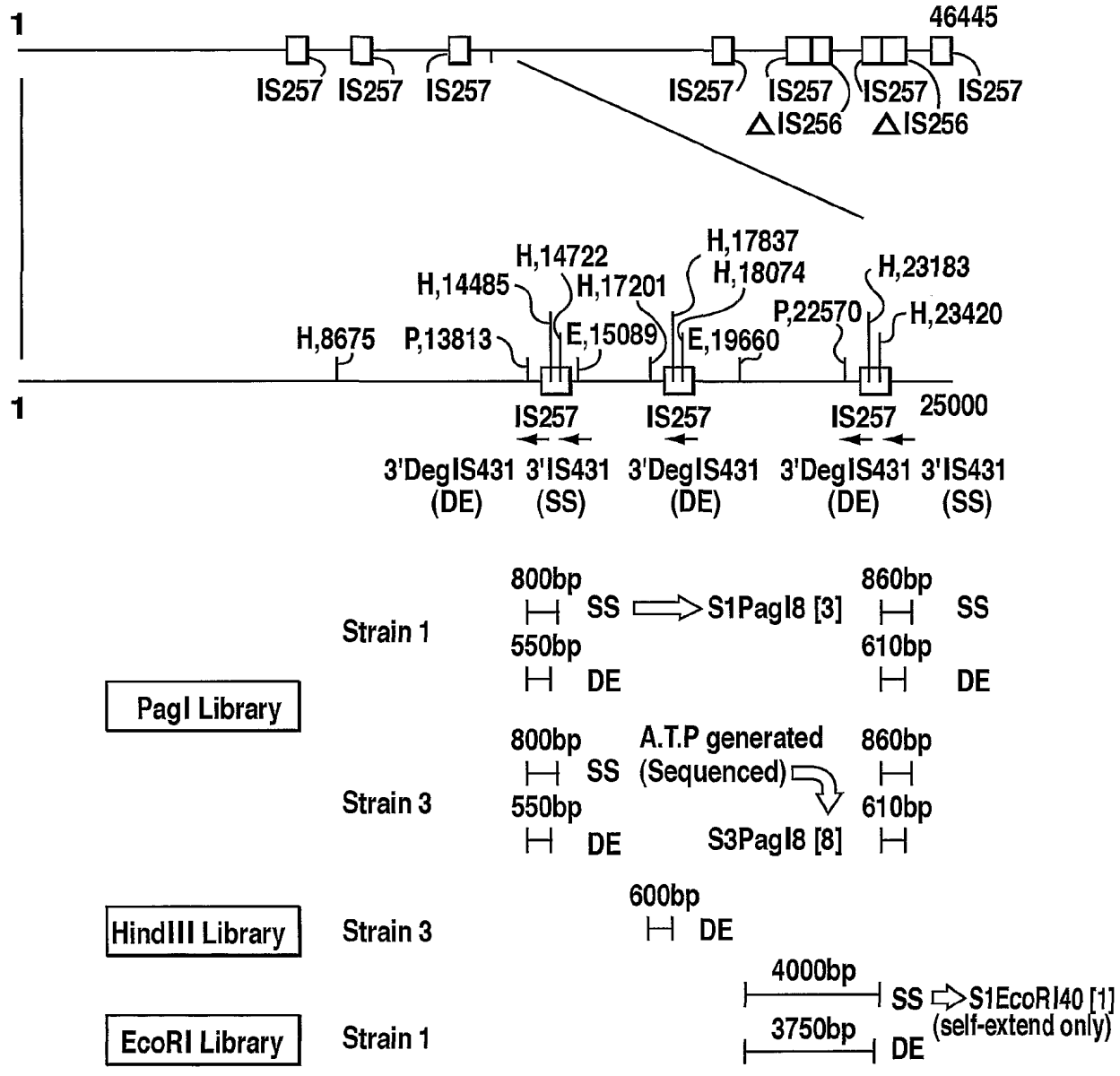

1
46445
IS257
IS257
IS257
IS257
IS257
IS257
IS257
ΔIS256
ΔIS256
H,14722
H,17837
H,14485
H,17201
H,18074
H,23183
P,22570
H,23420
H,8675
P,13813
E,15089
E,19660
1
IS257
IS257
IS257
25000
3'DegIS431 (DE)
3'IS431 (SS)
3'DegIS431 (DE)
3'DegIS431 (DE)
3'IS431 (SS)
PagI Library
Strain 1
800bp
SS
S1PagI8 [3]
860bp
SS
550bp
DE
610bp
DE
Strain 3
800bp
SS
A.T.P generated
(Sequenced)
860bp
550bp
610bp
DE
S3PagI8 [8]
HindIII Library
Strain 3
600bp
DE
EcoRI Library
Strain 1
4000bp
SS
S1EcoRI40 [1]
(self-extend only)
3750bp
DE

FIG. 9a

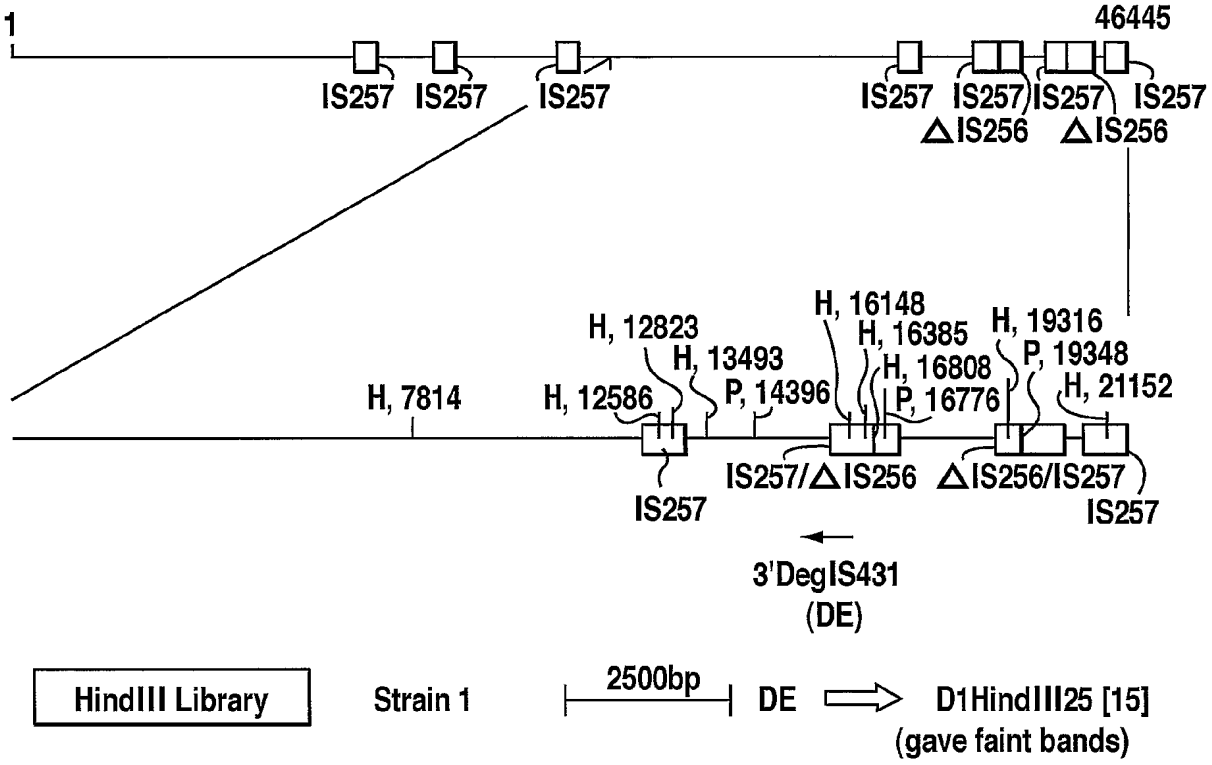

1
46445
IS257
IS257
IS257
IS257
IS257
IS257
IS257
ΔIS256
ΔIS256
H, 12823
H, 16148
H, 16385
H, 19316
H, 13493
P, 19348
H, 16808
H, 7814
H, 12586
P, 14396
P, 16776
H, 21152
IS257/ΔIS256
ΔIS256/IS257
IS257
IS257
3'DegIS431
(DE)
HindIII Library
Strain 1
2500bp
DE
D1HindIII25 [15]
(gave faint bands)

FIG. 9b

SEQUENCE LISTING

<110> MBI Fermentas Inc.

<120> Novel Method of Genomic Analysis

<130> 91642-7

<140>
<141> 2001-10-22

<150> US 09/695545
<151> 2000-10-23

<160> 27

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 1, 2, 3, 4, 5, 103
<223> n = A,T,C or G

<223> primer

<221> misc_feature
<222> 1, 2, 3, 4, 5, 103
<223> n = A,T,C or G

<221> misc_feature
<222> 1, 2, 3, 4, 5, 103
<223> n = A,T,C or G

<400> 1
nnnnnctcca gcggccgcgg tcgacgtacg atcagacgca taggtctccg gattttccgg 60
agacctagat agtctgcgtc gacagtcccg cggccgctgg agn 103

<210> 2
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 2
gctccagcgg ccgcggtcga cgtacgatca gacgcatagg tctccggatt ttccggagac 60
ctagatagtc tgcgtcgaca gtcccgcggc cgctggagct gca 103

<210> 3
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 3
aattcctcca gcggccgcgg tcgacgtacg atcagacgca taggtctccg gattttccgg 60
agacctagat agtctgcgtc gacagtcccg cggccgctgg agg 103

<210> 4
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 4
agcttctcca gcggccgcgg tcgacgtacg atcagacgca taggtctccg gattttccgg 60
agacctagat agtctgcgtc gacagtcccg cggccgctgg aga 103

<210> 5
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 5
catgactcca gcggccgcgg tcgacgtacg atcagacgca taggtctccg gattttccgg 60
agacctagat agtctgcgtc gacagtcccg cggccgctgg agt 103

<210> 6
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
gatagtctgc gtcgacagtc 20

<210> 7
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 7
caacgtgagc gatgtctagt 20

<210> 8
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 8
tgcagtcagc gcagactc 18

<210> 9
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 9
gccgcagatc gaagttac 18

<210> 10
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 10
gattatggat ggcaccactc 20

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 11
gttgaaggtg cctgatctgt 20

<210> 12
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 22
<223> n = A,T,C or G

<223> primer

<221> misc_feature
<222> 22
<223> n = A,T,C or G

<221> misc_feature
<222> 22
<223> n = A,T,C or G

<400> 12
tggtgcatat tcttgaaccc anckrta 27

<210> 13
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 13
gtagccgttg gctactatct aag 23

<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 27, 28
<223> n = A,T,C or G

<223> primer

<221> misc_feature
<222> 27, 28
<223> n = A,T,C or G

<221> misc_feature
<222> 27, 28
<223> n = A,T,C or G

<400> 14
caacgtgagc gatgtctagt ctgaagnn 28

<210> 15
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 15
cttcagacta gacatcgctc acgttg 26

<210> 16
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 16
ctgcagaagc atcttt 16

<210> 17
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 17
tcatgaatga taatca 16

<210> 18
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 18
aagcttgaaa tgagca 16

<210> 19
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 19
tcatgagtac tgatat 16

<210> 20
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 20
gaattcaccc acctca 16

<210> 21
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 21
tcatgataaa gaatct 16

<210> 22
<211> 56
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 22

gatagtctgc gtcgacagtc ccgcggccgc tggagctgca ctgcagaagc atcttt 56

<210> 23
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 23
gatagtctgc gtcgacagtc ccgcggccgc tggagttcat gaatgataat ca 52

<210> 24
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 24
gatagtctgc gtcgacagtc ccgcggccgc tggagaaagc ttgaaatgag ca 52

<210> 25
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 25
gatagtctgc gtcgacagtc ccgcggccgc tggagttcat gagtactgat at 52

<210> 26
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 26
gatagtctgc gtcgacagtc ccgcggccgc tggagggaat tcacccacct ca 52

<210> 27
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 27
gatagtctgc gtcgacagtc ccgcggccgc tggagttcat gataaagaat ct 52